

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2005, 01:35:43 ; Search time 7976 Seconds
(without alignments)

11002.056 Million cell updates/sec

Title: US-10-031-589-3

Perfect score: 1811

Sequence: 1 agtgcctgcggccctcgcgcg.....cggcgtgattgcgaacctcg 1811

Scoring table: IDENTITY NUC

Searched: 4708233 seqs, 2422767955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

GenEmbl: *
1: gb_ha: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_scs: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1811	100.0	1811	6	AX077432 Sequence
2	1811	100.0	1811	9	AK075358 Homo sapi
3	1631.4	90.1	1839	9	AY007158 Homo sapi
4	1621	89.5	1775	6	BD158346 Homo sapi
5	1621	89.5	1775	6	AX880378 Sequence
6	1621	89.5	1775	9	AK027733 Homo sapi
7	1597.6	88.2	1709	9	AK123762 Homo sapi
8	1434.8	79.2	1614	6	BD158376 Primer fo
9	1434.8	79.2	1614	6	AX880429 Sequence
10	1434.8	79.2	1614	6	AK027741 Homo sapi
11	1371	75.7	1407	9	BC022092 Homo sapi
12	854	47.2	199522	9	AC089999 Homo sapi
13	661.4	36.5	784	6	BD149092 Primer fo
14	661.4	36.5	784	6	AX869030 Sequence
15	562	31.0	597	6	BD149164 Primer fo
16	562	31.0	597	6	AX869102 Sequence
17	550	30.4	816	6	AX781130 Sequence
18	494	27.3	561	6	BD154159 Primer fo
19	494	27.3	561	6	AX874097 Sequence

20	472.8	26.1	1485	10	BC021365	BC021365 Mus muscu
21	465	25.7	602	6	CO729207	CO729207 Sequence
22	369	20.4	572	6	AX779761	AX779761 Sequence
23	353.8	19.5	263827	2	AC094579	AC094579 Rattus no
24	353.8	19.5	274026	2	AC129633	AC129633 Rattus no
25	341.8	18.9	578	6	BD154095	BD154095 Primer fo
26	341.8	18.9	578	6	AX874033	AX874033 Sequence
27	299.2	16.5	237928	2	AC137208	AC137208 Rattus no
28	273.2	15.1	221622	10	AC125183	AC125183 Mus muscu
29	241	13.3	199165	2	AC131195	AC131195 Rattus no
30	186.2	10.3	60319	2	AC102037	AC102037 Mus muscu
31	130.8	7.2	292	6	AX315440	AX315440 Sequence
32	126	7.0	60319	2	AC102037	AC102037 Mus muscu
33	125	6.9	244529	2	AC121112	AC121112 Mus muscu
34	82.8	4.6	125020	9	AF429315	AF429315 Homo sapi
35	61.6	3.4	125020	9	AF429315	AF429315 Homo sapi
36	60	3.3	60	6	CO535257	CO535257 Sequence
37	57.6	3.2	3153	5	BC070862	BC070862 Xenopus 1
38	56.8	3.1	174471	2	AC147859	AC147859 Macropus
39	56.6	3.1	141657	2	AC147760	AC147760 Macropus
40	56.6	3.1	163688	2	AC148274	AC148274 Macropus
41	55.2	3.0	1365	5	CR385236	CR385236 Gallus ga
42	54.4	3.0	24250	9	AY764389	AY764389 Homo sapi
43	54.4	3.0	39970	9	AP001167	AP001167 Homo sapi
44	54.4	3.0	146451	2	AC019309	AC019309 Homo sapi
45	54.4	3.0	185984	9	AP000942	AP000942 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AX077432 1811 bp DNA linear PAT 22-FEB-2001
DEFINITION Sequence 3 from Patent WO0107607.
ACCESSION AX077432
VERSION AX077432.1 GI:13121982
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
Ota, T., Iwaga, T., Nishikawa, T. and Kawai, Y.
Full length cDNA clones and proteins encoded thereby
Patent: WO 0107607-A 3 01-FEB-2001;
Helix Research Institute (JP)

FEATURES
source location/Qualifiers
1..1811
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
659..1468
/note="unnamed protein product"

CDS

/codon_start=1
/protein_id="CAC32398.1"
/db_xref="GI:13121983"
/translation="MKTPELVAGSMQTLQHRCRGGYRVKARTSYVDLTFGSPAG
TRPTPEPDPVWEKATRGVKEASALGAKSCETTPRGSGPTLPKKNYRP
ISHPPYCDSELFGSRSGASFGAPRMKGDALRLALMTPTPPRGSHGPRPEAP
LRATHPAGSTPEGPADDSOKLIMGGHSRRLKRGISHLTLNVSSTHPATSPAP
HTMGQDLRSTSGVTFRSPPLVTSRANVSISVSTPRRGATQKPKPMK"

ORIGIN

Query Match 100.0%; Score 1811; DB 6; Length 1811;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1811; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTGCCTGCGGCGCTTGGCGGCTTGGTACACACGACCTTGGTGTGGCAGAGAGAC 60
DB 1 AGTGCCTGCGGCGCTTGGCGGCTTGGTACACACGACCTTGGTGTGGCAGAGAGAC 60
QY 61 CCTTCCATGTTTAGAGGACCTCTGGGCTTACAGAGAGCTGGCGGCGGCGGCGGAGAC 120

```

Db      61  CCTCTCATGTTTAAAGGACCTCTGGGCTCAGAGAGGTGGCGCGCCCTCTGGGCGGAC 120
Qy      121  TCCCCCATCCGGGGGCGCGAATGTCCTGGGTCGCTCCGAGTGTCTGTGCTCTCC 180
Db      121  TCCCCCATCCGGGGGCGCGAATGTCCTGGGTCGCTCCGAGTGTCTGTGCTCTCC 180
Qy      181  TGGTGTGCGGGTGTCAAGTGTGGGTTCTGGGTTTCTGGAATTCGCGGGCGGTTTCAACGT 240
Db      181  TGGTGTGCGGGTGTCAAGTGTGGGTTCTGGGTTTCTGGAATTCGCGGGCGGTTTCAACGT 240
Qy      241  AGCCTGTGCGGCTCTCTGGGTAAGTCCGTCGCGCGCGGTGCGCCGCGAAGCGCTTAAAGC 300
Db      241  AGCCTGTGCGGCTCTCTGGGTAAGTCCGTCGCGCGCGGTGCGCCGCGAAGCGCTTAAAGC 300
Qy      301  TGGCGGGGGTTCGGGGGCGCGAATTCGGGCTGCGAGATTGACGGGGATTCGGGATGCA 360
Db      301  TGGCGGGGGTTCGGGGGCGCGAATTCGGGCTGCGAGATTGACGGGGATTCGGGATGCA 360
Qy      361  CCGGCGGCGCGCGCGCTCAGCGAGCGGGTCCAGACTGGTGGGGAAGAAGTGGCGGGAC 420
Db      361  CCGGCGGCGCGCGCGCTCAGCGAGCGGGTCCAGACTGGTGGGGAAGAAGTGGCGGGAC 420
Qy      421  GGGTCCCTGAGATCCCGATGCTTACGAGCGCAAGATCTGAGCTTTATAGGTGTGACTTA 480
Db      421  GGGTCCCTGAGATCCCGATGCTTACGAGCGCAAGATCTGAGCTTTATAGGTGTGACTTA 480
Qy      481  CACATGTGACTTCACTCAAGTTTGTGATCCGTAATAATGCAAAATTCGAAGCTACTTCA 540
Db      481  CACATGTGACTTCACTCAAGTTTGTGATCCGTAATAATGCAAAATTCGAAGCTACTTCA 540
Qy      541  CAGTGTGTTGAGAGATTAAATGAACAATGCTGTAAAGCTTTTGGACGAGGAGGCC 600
Db      541  CAGTGTGTTGAGAGATTAAATGAACAATGCTGTAAAGCTTTTGGACGAGGAGGCC 600
Qy      601  TCGGAACAGAGGCTTGGCGCGGAGAGACACTTGTCTCAACAGGAGCAACAGGACAGAT 660
Db      601  TCGGAACAGAGGCTTGGCGCGGAGAGACACTTGTCTCAACAGGAGCAACAGGACAGAT 660
Qy      661  GAAAGACCCCGTGGAGCTGGCCGTCAAGTGGGATGCAACCTCGGCTTCAAGCACCCCTG 720
Db      661  GAAAGACCCCGTGGAGCTGGCCGTCAAGTGGGATGCAACCTCGGCTTCAAGCACCCCTG 720
Qy      721  CCGAGGTGCTAACCGGGTCAAGGCGAGAGGTCAATATGAGGATGAGACTCTGTTTGGCAG 780
Db      721  CCGAGGTGCTAACCGGGTCAAGGCGAGAGGTCAATATGAGGATGAGACTCTGTTTGGCAG 780
Qy      781  CCGAGGAGGACCCGCGCTTACCCCAACCGGACTTTCGATCCGCTGGGTGAGAGAGCTTAA 840
Db      781  CCGAGGAGGACCCGCGCTTACCCCAACCGGACTTTCGATCCGCTGGGTGAGAGAGCTTAA 840
Qy      841  CAGAACCAAGAGGCTGGGCAAGAGGACATGCAAGGCTTGGGGCAAAAGGAGCTGTGA 900
Db      841  CAGAACCAAGAGGCTGGGCAAGAGGACATGCAAGGCTTGGGGCAAAAGGAGCTGTGA 900
Qy      901  GACCAACCCCTCAAGGGGAGCAACCCCAACCTTCAACAAGGAAGAACAATATCAG 960
Db      901  GACCAACCCCTCAAGGGGAGCAACCCCAACCTTCAACAAGGAAGAACAATATCAG 960
Qy      961  ACCCATGAGCACAACCCGCTTACTGTGATGATCGCTGTTGGCTCCGATCTGAAGG 1020
Db      961  ACCCATGAGCACAACCCGCTTACTGTGATGATCGCTGTTGGCTCCGATCTGAAGG 1020
Qy      1021  CGCCAGCTTGGGGGCGCGGGGATGCGAAGGGGGATGCGCAAAAGCTCGTGTCTCTT 1080
Db      1021  CGCCAGCTTGGGGGCGCGGGGATGCGAAGGGGGATGCGCAAAAGCTCGTGTCTCTT 1080
Qy      1081  GTGAGCGCCACCACTTACCCCAAGGGGATGACACTTCCCGCCCAAGGAGGACCACT 1140
Db      1081  GTGAGCGCCACCACTTACCCCAAGGGGATGACACTTCCCGCCCAAGGAGGACCACT 1140
Qy      1141  GCGAGCATTCACCCAGCTGGTCCCTCAAGAGACAGACCGGGGCGACGCGAGACTCCA 1200

```

```

Db      1141  GCGAGCATTCACCCAGCTGGTCCCTCAAGAGACAGACCGGGGCGACGCGAGACTCCA 1200
Qy      1201  GAAGTTATCTAATGGGTGGTTTACACTTTCAGGCCCCCTGAAGCGGGGACTTTCCGATTTC 1260
Db      1201  GAAGTTATCTAATGGGTGGTTTACACTTTCAGGCCCCCTGAAGCGGGGACTTTCCGATTTC 1260
Qy      1261  CCTCACCCACTGAATGTCCTCCAGCACTGGTCAATCCAGCCACCAAGTCCGCCACAA 1320
Db      1261  CCTCACCCACTGAATGTCCTCCAGCACTGGTCAATCCAGCCACCAAGTCCGCCACAA 1320
Qy      1321  TGGGCTCAGAGATCTCAGGCTTTCAGCTCAGGGGTGACCTTCGGAAGCCCTCGTGTAC 1380
Db      1321  TGGGCTCAGAGATCTCAGGCTTTCAGCTCAGGGGTGACCTTCGGAAGCCCTCGTGTAC 1380
Qy      1381  TTTCCAGGGCTGCTCAGTTAGCAATTCAGTGCATCAATCCCAAGAGGTGGGGCAC 1440
Db      1381  TTTCCAGGGCTGCTCAGTTAGCAATTCAGTGCATCAATCCCAAGAGGTGGGGCAC 1440
Qy      1441  CCAAGAACCAAGACCCCTTGAATGATCTTTCATCAGAGGTTGCTTATGGGGCAC 1500
Db      1441  CCAAGAACCAAGACCCCTTGAATGATCTTTCATCAGAGGTTGCTTATGGGGCAC 1500
Qy      1501  GCGGACAGTATGAGCCCTTTCAGAGGTGAGAGACATTCATCAACCAAGAACCCAGG 1560
Db      1501  GCGGACAGTATGAGCCCTTTCAGAGGTGAGAGACATTCATCAACCAAGAACCCAGG 1560
Qy      1561  TATTAAAGAACCCCTGTTGGGGGAGACAGACATCAGAGGGGTGGGAGGCTCCCTTT 1620
Db      1561  TATTAAAGAACCCCTGTTGGGGGAGACAGACATCAGAGGGGTGGGAGGCTCCCTTT 1620
Qy      1621  ATCCTGACATCTCTAGTGCATCTTTCCTTTCCTCCGATTCGCGATTGGGGGCGAC 1680
Db      1621  ATCCTGACATCTCTAGTGCATCTTTCCTTTCCTCCGATTCGCGATTGGGGGCGAC 1680
Qy      1681  CTCTAAGATGCTCTCTCCAGCCCTGTCTCAACCAATCTCAATTAATGTCGAACCCAGG 1740
Db      1681  CTCTAAGATGCTCTCTCCAGCCCTGTCTCAACCAATCTCAATTAATGTCGAACCCAGG 1740
Qy      1741  GGCCTGGACCTCCCAATCATTCATTTGCTGTCGCAAGTGCAGATTAACGGCGTAT 1800
Db      1741  GGCCTGGACCTCCCAATCATTCATTTGCTGTCGCAAGTGCAGATTAACGGCGTAT 1800
Qy      1801  TGCCAACTGG 1811
Db      1801  TGCCAACTGG 1811

RESULT 2
AK075358
LOCUS      1811 bp      mRNA      linear      PRI 03-SEP-2002
DEFINITION Homo sapiens cDNA PSEC0043 fls, clone NTZRP1001002.
ACCESSION AK075358
VERSION    AK075358.1 GI:22761393
KEYWORDS   oligo capping; fls (full insert sequence) .
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1
            Oca,T., Nishikawa,T., Suzuki,Y., Kawai-Hio,Y., Hayashi,K.,
            Ishii,S., Saito,K., Yamamoto,J., Wakamatsu,A., Nagai,T.,
            Nakamura,Y., Nagahari,K., Sugano,S. and Isogai,T.,
            HRI human cDNA sequencing project
            Unpublished
            2 (bases 1 to 1811)
            Isogai,T. and Yamamoto,J.,
            Direct Submission
            Submitted (20-MAR-2002) Takao Isogai, Helix Research Institute,
            Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
            HRI human cDNA sequencing project; cDNA 5'- & 3'-end one pass
            sequencing, clone selection and full insert sequencing; Helix
            Research Institute (supported by Japan Key Technology Center etc.);

```


QY 1741 GGCTGGACACTCCACATCATTCATGTGCTGGCAAGTGGAATTAACGGCGTAT 1800
DB 1741 GGCTGGACACTCCACATCATTCATGTGCTGGCAAGTGGAATTAACGGCGTAT 1800
QY 1801 TGCCACCTGG 1811
DB 1801 TGCCACCTGG 1811

RESULT 3
AY007158 1839 bp mRNA linear PRI 31-AUG-2000
LOCUS Homo sapiens clone CDABP0113 mRNA sequence.
DEFINITION AY007158
ACCESSION AY007158.1 GI:9956071
VERSION F11 CDNA.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1839) Anderson, B., Worley, K.C., Muzny, D.M., Ding, Y., Liu, W.,
Andersson, B., Wentland, M.A., Ricafente, J.Y., Liu, W. and Gibbs, R.A.
TITLE 'A double adaptor' method for improved shotgun library construction
JOURNAL Anal. Biochem. 236 (1), 107-113 (1996)
MEDLINE 96207227
PUBMED 8619474

REFERENCE 2 (bases 1 to 1839) Yu, W., Andersson, B., Worley, K.C., Muzny, D.M., Ding, Y., Liu, W.,
Ricafente, J.Y., Wentland, M.A., Lennom, G. and Gibbs, R.A.
TITLE Large-scale concatenation cDNA sequencing
JOURNAL Genome Res. 7 (4), 353-358 (1997)
MEDLINE 97264341
PUBMED 9110174

REFERENCE 3 (bases 1 to 1839) Zhou, J., Yu, W., Tang, H., Mei, G., Tsang, Y.T.M., Bouck, J., Gibbs, R.A.
AUTHORS and Margolin, J.F.
TITLE Direct Submission
JOURNAL Children's Cancer Center, Baylor College of Medicine, Houston, TX
77030, USA
COMMENT The clone request should be directed to Dr. J. Margolin at
Pediatrics-Hematology & Oncology, Texas Children's PEIGIN Center
10251, Houston, Texas 77030, USA. Telephone: 713-770-4583
margolinbcm.tmc.edu.

FEATURES
source
Location/Qualifiers
1..1839
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CDABP0113"
/clone_lib="constructed by Y.T.M. Tsang"
/dev_stage="Infant"
/note="from patient with acute lymphoblastic leukemia"

ORIGIN
Query Match 90.1%; Score 1631.4; DB 9; Length 1839;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1632; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 179 CTTGGTGGCTGGGCAAGTGTGAGTTCGAGTTTCGATTCGCGGCGCTTCAAC 238
DB 1 CTTGGTGGCTGGGCAAGTGTGAGTTCGAGTTTCGATTCGCGGCGCTTCAAC 60
QY 239 GTACCTGTGCGGCTCTCTCGGAGTCCGTCGCGCGCGGCTCCCGGAGCGGCTTAG 298
DB 61 GTACCTGTGCGGCTCTCTCGGAGTCCGTCGCGCGCGGCTCCCGGAGCGGCTTAG 120
QY 299 GCTGCGCGGCTCTCTCGGAGTCCGTCGCGCGCGGCTCCCGGAGCGGCTTAG 358
DB 121 GCTGCGCGGCTCTCTCGGAGTCCGTCGCGCGCGGCTCCCGGAGCGGCTTAG 180
QY 359 CACGCGCGGCTCTCTCGGAGTCCGTCGCGCGCGGCTCCCGGAGCGGCTTAG 418
DB 1261 CACGCGCGGCTCTCTCGGAGTCCGTCGCGCGCGGCTCCCGGAGCGGCTTAG 1320

DB 181 CACGCGCGGCTCTCTCGGAGTCCGTCGCGCGGCTCCCGGAGCGGCTTAG 240
QY 419 AGCGGCTCTCTCGGAGTCCGTCGCGCGGCTCCCGGAGCGGCTTAG 478
DB 241 AGCGGCTCTCTCGGAGTCCGTCGCGCGGCTCCCGGAGCGGCTTAG 300
QY 479 TACACATGTGACTTCACTCAGTGTGATCCGTAATTAATGAATTCGAGTACTT 538
DB 301 TACACATGTGACTTCACTCAGTGTGATCCGTAATTAATGAATTCGAGTACTT 360
QY 539 CACAGTGTGAGAGATTAATGAATTCGTAATTAATGAATTCGTAATTCGTAAT 598
DB 361 CACAGTGTGAGAGATTAATGAATTCGTAATTAATGAATTCGTAATTCGTAAT 420
QY 599 CTTGGAGAGAGGCTCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 658
DB 421 CTTGGAGAGAGGCTCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 659 ATGAAGAGAGGCTCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 718
DB 481 ATGAAGAGAGGCTCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
QY 719 TGCCAGAGTGTACCGGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 778
DB 541 TGCCAGAGTGTACCGGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 779 AGCCGAG 838
DB 601 AGCCGAG 660
QY 839 AACAG 898
DB 661 AACAG 720
QY 899 GAG 958
DB 721 GAG 780
QY 959 AG 1018
DB 781 AG 840
QY 1019 GAG 1078
DB 841 GAG 900
QY 1079 TTGTGAG 1138
DB 901 TTGTGAG 960
QY 1139 CTGCGAG 1198
DB 961 CTGCGAG 1020
QY 1199 CAG 1258
DB 1021 CAG 1080
QY 1259 TCCCTCAG 1318
DB 1081 TCCCTCAG 1140
QY 1319 AATGAG 1378
DB 1141 AATGAG 1200
QY 1379 ACTTCCAG 1438
DB 1201 ACTTCCAG 1260
QY 1439 ACCAG 1498
DB 1261 ACCAG 1320

Qy	1499	ACGGGAGAGGTATGCCCCCTTGCCAGGGTAGAGGAATTCATCACCAGGAAACCCA	1558
Db	1321	ACGGGAGAGGTATGCCCCCTTGCCAGGGTAGAGGAATTCATCACCAGGAAACCCA	1380
Qy	1559	GGATATTAAAGAACCCCTGTGGGGGACACAGACATAGCAGGGGGGAGTGCCTCCT	1618
Db	1381	GGATATTAAAGAACCCCTGTGGGGGACACAGACATAGCAGGGGGGAGTGCCTCCT	1440
Qy	1619	TTATTCGTACAATCTCTAGTCGATTCCTTGCCGATTGCGAATTTGGGGGCC	1678
Db	1441	TTATTCGTACAATCTCTAGTCGATTCCTTGCCGATTGCGAATTTGGGGGCC	1500
Qy	1679	ACCTCTAAGATGCTCTCTCCAGCCCTGTCTCAACACATATCTCAAATTAAGTGCACACCA	1738
Db	1501	ACCTCTAAGATGCTCTCTCCAGCCCTGTCTCAACACATATCTCAAATTAAGTGCACACCA	1560
Qy	1739	GGGGCGTGGGCACTCCCAATTCATTCATTTGTCTTGCTGCGCAAGTGCGAATTAAGCGGTG	1798
Db	1561	GGGGCGTGGGCACTCCCAATTCATTCATTTGTCTTGCTGCGCAAGTGCGAATTAAGCGGTG	1620
Qy	1799	ATTGCGAACTGG 1811	
Db	1621	ATTGCGAACTGG 1633	

LOCUS	BD158346	1775 bp	DNA linear	PAT 17-JAN-2003
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.			
ACCESSION	BD158346			
VERSION	BD158346.1	GI:27864104		
KEYWORDS	JP 2002191363-A/13189.			
SOURCE	Homo sapiens (human)			

REFERENCE
1 (bases 1 to 1775)
Ota, T., Isegai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 13189 09-JUL-2002;

COMMENT

PN JP 2002191363-A/13189
PD 09-JUL-2002
PI 28-JUL-2000 JP 2000280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
PI SAITO,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUJI OTSUKI
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10,
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT CDS (469) .. (1275) .

FEATURES	
SOURCE	Location/Qualifiers
1. 1775	
/organism="Homo sapiens"	
/mol_type="Genomic DNA"	
/db_xref="taxon:9606"	

Query Match	89.5%	Score 1621	DB 6	Length 1775
Best Local Similarity	100.0%	Prod. No. 0		
Matches 1621	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	191	GTGCAGAAAGTGTGGGATTCTCGAGATTCGGGAGCCGTGTACACGTAGCCTGTGCC	250	
Db	1	GTGCAGAAATGTCTGGGATTCGTGGATTCTCGAGATTCGGGAGCCGTGTACACGTAGCCTGTGCC	60	

OY	251	GGCTCTCGGGGAGATCCGTCGCGCGGGTCCCGGGACGGGCTTACGGCTCGCGGGGT	310
Db	61	GGCTCTCGGGAGATCCGTCCGCGCGGGTCCCGGGACGGGCTTACGGCTCGCGGGGT	120
OY	311	CGGGGGCCCGCAGGCATTCCGGGCTGCAGATTGACGGGGATCCCGGATGACCGCGGCC	370
Db	121	CGGGGGCCCGCAGGCATTCCGGGCTGCAGATTGACGGGGATCCCGGATGACCGCGGCC	180
OY	371	CGCGCCCTCACCGACGGTCCGACTGTGTGGAAAGAAAGTTCGGGGACGGGTCCCTGA	430
Db	181	CGCGCCCTCACCGACGGTCCGACTGTGTGGAAAGAAAGTTCGGGGACGGGTCCCTGA	240
OY	431	GGATCCCGATGTCTACGAGCAGAAAGTGCACACTTATAGGTGACCTTACATGTGAC	490
Db	241	GGATCCCGATGTCTACGAGCAGAAAGTGCACACTTATAGGTGACCTTACATGTGAC	300
OY	491	TTCACTCAGTTTGTGATCCGTAAATGACAAATTCGAACTACTTCACAGTGTGT	550
Db	301	TTCACTCAGTTTGTGATCCGTAAATGACAAATTCGAACTACTTCACAGTGTGT	360
OY	551	GAGAGATTAAATGAAACAATGCTTTGTAAGCTCTTTGACGAGGAGAGCTTCGGAAGCAG	610
Db	361	GAGAGATTAAATGAAACAATGCTTTGTAAGCTCTTTGACGAGGAGAGCTTCGGAAGCAG	420
OY	611	GGCTTCGGCGGACAGACACTGTGTGACCAAGGACCAACGGACGACTGAAGACCCCC	670
Db	421	GGCTTCGGCGGACAGACACTGTGTGACCAAGGACCAACGGACGACTGAAGACCCCC	480
OY	671	GTGAGCTGCGCGTACGTGGGATGAGAGACCTTCGCGCTTTCACACCGCTGCGGAGTGGC	730
Db	481	GTGAGCTGCGCGTACGTGGGATGAGAGACCTTCGCGCTTTCACACCGCTGCGGAGTGGC	540
OY	731	TACCGGGTCAAAGCCAGGACGTCAATGTGAGATGAGACTTGTTTGGCAGCCAGAGGC	790
Db	541	TACCGGGTCAAAGCCAGGACGTCAATGTGAGATGAGACTTGTTTGGCAGCCAGAGGC	600
OY	791	ACCGGGCTTACCCCAACCGGACTTGCATCCGCTTCGGGTGAGAGAGCTTACAGAACAGA	850
Db	601	ACCGGGCTTACCCCAACCGGACTTGCATCCGCTTCGGGTGAGAGAGCTTACAGAACAGA	660
OY	851	GGCGTGGGCAAGAGGAGCATTCGAAAGGCTTGGGGGGCAAGGGGAGCTGTGAGACCAACCCC	910
Db	661	GGCGTGGGCAAGAGGAGCATTCGAAAGGCTTGGGGGGCAAGGGGAGCTGTGAGACCAACCCC	720
OY	911	TCAAAGGGCAGACCCGCCCTCAACAAGAAAGAAACAAATACAGACCATCAGC	970
Db	721	TCAAAGGGCAGACCCGCCCTCAACAAGAAAGAAACAAATACAGACCATCAGC	780
OY	971	CACACCCCGTCTTACTGTATGATGTGCTGTGTTGCTCCCGATCTGAAGGGCCAGCTTC	103
Db	781	CACACCCCGTCTTACTGTATGATGTGCTGTGTTGCTCCCGATCTGAAGGGCCAGCTTC	840
OY	1031	GGGGCCCCGGAGATGGCGAAGGGGGATGCCCGAAGCTCCGCTCTCTGTGGAGCGCA	109
Db	841	GGGGCCCCGGAGATGGCGAAGGGGGATGCCCGAAGCTCCGCTCTCTGTGGAGCGCA	900
OY	1091	CCACCTACCCCGAAGGGTATGACCATCGCGCCCGCCCAAGGAGGACCACTGTGAGACATT	115
Db	901	CCACCTACCCCGAAGGGTATGACCATCGCGCCCGCCCAAGGAGGACCACTGTGAGACATT	960
OY	1151	CACCCAGCTGTGCTCTCCAGAAGAGAGCCGGGGCCAGCGGACATCTCCAGAAATTATCT	121
Db	961	CACCCAGCTGTGCTCTCCAGAAGAGAGCCGGGGCCAGCGGACATCTCCAGAAATTATCT	102
OY	1211	ATGGGTGGGTACACTCTTTCACGGCCCCCTGAAGCGGGGACCTTCCCATTTCCCTCACCCAC	127
Db	1021	ATGGGTGGGTACACTCTTTCACGGCCCCCTGAAGCGGGGACCTTCCCATTTCCCTCACCCAC	108
OY	1271	CTGAATGTCCCGACGACTGTGTATCAGCCCAACAGTGTCCCGCACACAATAATGGGCTCAG	133
Db	1081	CTGAATGTCCCGACGACTGTGTATCAGCCCAACAGTGTCCCGCACACAATAATGGGCTCAG	114
OY	1331	GATCTCAGGGCTTTCACGTCAGGGGTGACTTTCGGAGCCCCCTGTGTGACTTTCAGGGCT	139

QY	191	GTGGAAGTCTGGGTTCTGGATTTCGATTCGGCGGCGCGTTCAACGTAAGCTGTGCC	250
Db	1	GTGGAAGTCTGGGTTCTGGATTTCGATTCGGCGGCGCGTTCAACGTAAGCTGTGCC	60
QY	251	GGCTCCTCGGGTAGTCCGTCGCGCGCGGTGCCCGGACCGGCTTAGGCTCCGGGGGT	310
Db	61	GGCTCCTCGGGTAGTCCGTCGCGCGCGGTGCCCGGACCGGCTTAGGCTCCGGGGGT	120
QY	311	CCGGGGCCCCAGGCAATTCGGGCTGCGAGATTGACGGGGATTCGCGATGCAACCGGCGCCC	370
Db	121	CCGGGGCCCCAGGCAATTCGGGCTGCGAGATTGACGGGGATTCGCGATGCAACCGGCGCCC	180
QY	371	CCGGGCCCTCAACGACCGGGTCCAGACCTGGGGGAAAGAGGTGGGGGACCGGGTCCCTGA	430
Db	181	CCGGGCCCTCAACGACCGGGTCCAGACCTGGGGGAAAGAGGTGGGGGACCGGGTCCCTGA	240
QY	431	GGATCCCGATGCTTACGAGCCAAAGATCTCAGCTTTATAGTGTGACCTTACACATGTGAC	490
Db	241	GGATCCCGATGCTTACGAGCCAAAGATCTCAGCTTTATAGTGTGACCTTACACATGTGAC	300
QY	491	TTCACTCAAGTTTGTGTATCCGTAAATGGAACAATTCGAAGCTACTTCAACAGTGGTT	550
Db	301	TTCACTCAAGTTTGTGTATCCGTAAATGGAACAATTCGAAGCTACTTCAACAGTGGTT	360
QY	551	GAGAGATTAAATGAACAATGCTTTGTAAGCTCTTTGCAGAGGGAGCCTCGGAACAG	610
Db	361	GAGAGATTAAATGAACAATGCTTTGTAAGCTCTTTGCAGAGGGAGCCTCGGAACAG	420
QY	611	GGCTGGCCGGCAGAGCAACTGTCTTCAACAGGACCAACAGGAGCATGAAAGACCCCC	670
Db	421	GGCTGGCCGGCAGAGCAACTGTCTTCAACAGGACCAACAGGAGCATGAAAGACCCCC	480
QY	671	GTGAGTGTGCGCTCAAGGGAATGACAACCTGGGCTTTAGACACCGCTGGCCGAGTGGC	730

QY	731	TACCGGGTCAAGGCGAAGACGTCAATATGTGATAGACTCTGTTTGGCAGGCCAGACGC	790
Db	541	TACCGGGTCAAGGCGAAGACGTCAATATGTGATAGACTCTGTTTGGCAGGCCAGACGC	600
QY	791	ACCCGGCTTACCCCAACGGACTTTCGATCCGCGCTTCGGGTGGAGAAGGCTTAACAGAACAGA	850
Db	601	ACCCGGCTTACCCCAACGGACTTTCGATCCGCGCTTCGGGTGGAGAAGGCTTAACAGAACAGA	660
QY	851	GGCGTGGGCAAGAGGAGCATCGAAGGCGCTTTGGGGCAAGGGAGACTGTGAGACCAACCCCC	910
Db	661	GGCGTGGGCAAGAGGAGCATCGAAGGCGCTTTGGGGCAAGGGAGACTGTGAGACCAACCCCC	720
QY	911	TCMAAGGGGCAAGACCCCAACCTTCACACCAAGAAAGAAAGAAACAAATACAGCCCATCAGC	970
Db	721	TCMAAGGGGCAAGACCCCAACCTTCACACCAAGAAAGAAAGAAACAAATACAGCCCATCAGC	780
QY	971	CACACCCCGTCTTAAGTGTGATGAGTCCGCTGTTTGAGTCCCGATCTGGAAGGGCGCAGCTTC	1030
Db	781	CACACCCCGTCTTAAGTGTGATGAGTCCGCTGTTTGAGTCCCGATCTGGAAGGGCGCAGCTTC	840
QY	1031	GGGGCCCCGGGAGATGGGGAAGGGGGAGACCGCAAAAGTCCGTGTCTCTTGTGGACGCCA	1090
Db	841	GGGGCCCCGGGAGATGGGGAAGGGGGAGACCGCAAAAGTCCGTGTCTCTTGTGGACGCCA	900
QY	1091	CCACTTACCCCCAGAGGTAGCCACTTCGCCCCCGCCCCAGGAGGACCACTTCGAGCCATT	1150
Db	901	CCACTTACCCCCAGAGGTAGCCACTTCGCCCCCGCCCCAGGAGGACCACTTCGAGCCATT	960
QY	1151	CACCCACCTGTCCCTCCAAAGACAGAGCCGAGGCGCAGCGCAGACTCCCGAATATTATCT	1210
Db	961	CACCCACCTGTCCCTCCAAAGACAGAGCCGAGGCGCAGCGCAGACTCCCGAATATTATCT	1020

Db 1021 ATGGGTGGGTACACTCTTCACGCCCCCTGAAGCGGGGACTTTCCTCCCATTTCCCTACCCAC 1080

Oy		1271	CTGAATGTCGCCAGACACTGTGCATCTCACCAACAAGTGGCCCCCACAACAATAATGGGCGTTAG	1330
Db		1081	CTGAATGTCGCCAGACACTGTGCATCTCACCAACAAGTGGCCCCCACAACAATAATGGGCGTTAG	1140
Oy		1331	GATCTCAGGCGCTTCCACGCTCAGAGGGTGAACTTTCGGAGCCCCCTGTGTAATTCTTCAGAGGCT	1390
Db		1141	GATCTCAGGCGCTTCCACGCTCAGAGGGTGAACTTTCGGAGCCCCCTGTGTAATTCTTCAGAGGCT	1200
Oy		1391	CGCTTCAGTTAGCATTTTCAGTGCATCTAACCCCAAGAAGTGGGGCCACCACGAAGAACCA	1450
Db		1201	CGCTTCAGTTAGCATTTTCAGTGCATCTAACCCCAAGAAGTGGGGCCACCACGAAGAACCA	1260
Oy		1451	AAGCCCCCTTGGAAATGATTAATCTCTTTCATCAAGGGGTGGCTPAATGGGGGCCAGAGGCAAGGT	1510
Db		1261	AAGCCCCCTTGGAAATGATTAATCTCTTTCATCAAGGGGTGGCTPAATGGGGGCCAGAGGCAAGGT	1320
Oy		1511	ATGGCCCCCTTGCACAGGAGTAGAAGAGCAATTCAATCAACCCAGGGAACCCCAGTATTAAAGAA	1570
Db		1321	ATGGCCCCCTTGCACAGGAGTAGAAGAGCAATTCAATCAACCCAGGGAACCCCAGTATTAAAGAA	1380
Oy		1571	GCCCCCTGTGGGGGCGACAGACATATGCAAGGGGTGGGCAGTGCTTCCTTTATCTTGACAA	1630
Db		1381	GCCCCCTGTGGGGGCGCGACAGACATATGCAAGGGGTGGGCAGTGCTTCCTTTATCTTGACAA	1440
Oy		1631	TCTCTAGTCGATTCCTTGGCTTTTTCTCCGAAATGGAGATTGGGGGGCCACCTTAAGATG	1690
Db		1441	TCTCTAGTCGATTCCTTGGCTTTTTCTCCGAAATGGAGATTGGGGGGCCACCTTAAGATG	1500
Oy		1691	CCTCTCTTCACAGCCCTGTCTCAACCAATATCCAATAATTAAGTGCACCAACCCAGGGGCGTTGGCAC	1750
Db		1501	CCTCTCTTCACAGCCCTGTCTCAACCAATATCCAATAATTAAGTGCACCAACCCAGGGGCGTTGGCAC	1560
Oy		1751	CTCCCAATCAATCAATTCATTTGTTCTGCTGCGCAAGTGCGAATPAAACGGCGTGATTTGCCAACCTG	1810
Db		1561	CTCCCAATCAATCAATTCATTTGTTCTGCTGCGCAAGTGCGAATPAAACGGCGTGATTTGCCAACCTG	1620
Oy		1811 G 1811		
Db		1621 G 1621		
RESULT 6 AK027733	LOCUS	AK027733	1775 bp	mRNA linear PRI 30-JAN-2004
DEFINITION	Homo sapiens cDNA FLJ14827 fis, clone OVARCI000886.			
ACCESSION	AK027733			
VERSION	AK027733.1 GI:14042629			
KEYWORDS	oligo capping; fis (full insert sequence).			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE				
AUTHORS	Ura,T., Suzuki,Y., Nishikawa,T., Otsubaki,T., Sugiyama,T., Irie,R., Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H., Sekine,M., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y., Nagabari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuna,M., Shiratori,A., Sudo,H., Hosoi,T., Kahu,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Puruya,T., Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M., Niinomiya,K., Ishibashi,T., Yamashta,H., Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M., Hiraoka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S., Yosida,M., Hottur,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A., Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K., Arita,M., Imose,N., Mutsashino,K., Yuuki,H., Oshima,A., Saeaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T., Shohate,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S., Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H., Goto,Y., Shimizu,F., Wakebe,H., Hisigaki,H., Watanabe,T., Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M.,			

Ozeki, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,
Izumi, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
Inagaki, H., Ikema, Y., Okamoto, S., Ohtani, R., Kawakami, T.,
Noguchi, S., Itoh, T., Shigeoka, K., Senba, T., Matsumura, K.,
Oyama, M., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,
Sato, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,
Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,
Nagase, T., Nomura, N., Kikuchi, H., Masuh, Y., Yamashita, R.,
Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs

JOURNAL
PUBMED
REFERENCES
AUTHORS

TITLE

JOURNAL
PUBMED
REFERENCES
AUTHORS

TITLE

JOURNAL
PUBMED
REFERENCES
AUTHORS

TITLE

JOURNAL
PUBMED
REFERENCES
AUTHORS

COMMENT

FEATURES
SOURCE

CDS

ORIGIN

Query Match 89.5%; Score 1621; DB 9; Length 1775;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1621; Conservative 0; Mismatches 0; Indels 0; Gaps 0

191 GTGCAAGTGTGGTTCTGGATTTCGGATTCGGCGGCCGTTACACAGTAGCTGTGC 250
Db 1 GTGCAAGTGTGGTTCTGGATTTCGGATTCGGCGGCCGTTACACAGTAGCTGTGC 60

QY 251 GGCTCTCGGGTAGTGCCGTCCGCGCGCGGTGCCCCGGGACGGCCTTAGGCTGCCGGGCT 310
Db 61 GGCTCTCGGGTAGTGCCGTCCGCGCGCGGTGCCCCGGGACGGCCTTAGGCTGCCGGGCT 120

QY 311 CCGGAGGCCAAGCATTCGGGCTGCGAGATTGAACGGGAGATCCGGAGATGACCGCGCGCC 370
Db 121 CCGGAGGCCAAGCATTCGGGCTGCGAGATTGAACGGGAGATCCGGAGATGACCGCGCGCC 180

Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986
NEO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing;
Research Association for Biotechnology; cDNA library construction;
5' - 3' end one pass sequencing and clone selection; Helix
Research Institute (supported by Japan Key Technology Center etc.)
and Department of Virology, Institute of Medical Science,
University of Tokyo

Location/Qualifiers

1..1775

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="OVARC1000886"
/cissue_type="ovary, tumor tissue"
/clone_id="OVARC1"
/note="Cloning vector: pME18SF33"
469..1278
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB55328.1"
/db_xref="GI:14042630"
/translation="MKTPIVELAVSGMOTLGLQHRGRGVRYKARTSYDETLFGSPAPA
TRPPDFDPDPEVWEKANRTGREGKEASALGADSKSCETTPRGSPPTLPKRNKYRIR
ISHRPYCDSILFGRSEBGSFGAPRAAKGAAALRALMLPPPRPSHSRPREAILRAHHP
GPSTEBEGPADSDSKLMGSLHSRRPLKRGISHLTHLVNSTGHPTSALHTNPDLRLBST
SVTPFRSPVLTISPARVSIVSPTEPRGATGPKPKPMK"

QY 371 CCGGCGCCCTCAACGAGCGGATCAGACCTGAGGAGAGAGGATGCGGGGACGGGATCCCTGA 430
 Db 181 CCGGCGCCCTCAACGAGCGGATCAGACCTGAGGAGAGAGGATGCGGGGACGGGATCCCTGA 240
 QY 431 GGATCCGATGCTCAGACGAGCCAGATGCTCAGCTTTAAGGTGATCACAATGTGAC 490
 Db 241 GGATCCGATGCTCAGACGAGCCAGATGCTCAGCTTTAAGGTGATCACAATGTGAC 300
 QY 491 TTCACTCAGTTTGTATCCGTAAATGACAAATTCGAAAGCTACTTTCACAGTGTGTT 550
 Db 301 TTCACTCAGTTTGTATCCGTAAATGACAAATTCGAAAGCTACTTTCACAGTGTGTT 360
 QY 551 GAGGAGATTAAATGAATGCTTTGTAAGCTCTTTGACAGAGAGAGCTCGGAAGCAG 610
 Db 361 GAGGAGATTAAATGAATGCTTTGTAAGCTCTTTGACAGAGAGAGCTCGGAAGCAG 420
 QY 611 GGCCTGCGCGGACAGACACCTGCTGACACAGGACCAACAGGACATGAAACCCCG 670
 Db 421 GGCCTGCGCGGACAGACACCTGCTGACACAGGACCAACAGGACATGAAACCCCG 480
 QY 671 GTGAGAGTGGCCCTCAGTGGGATGACACCTCGGCTTACAGACCGGCTGCGAGGTGGC 730
 Db 481 GTGAGAGTGGCCCTCAGTGGGATGACACCTCGGCTTACAGACCGGCTGCGAGGTGGC 540
 QY 731 TACCGGGTCAAGGCGCAGACGTCATATGTGATGAGACTGTTTGGCAGCGCCAGCAGC 790
 Db 541 TACCGGGTCAAGGCGCAGACGTCATATGTGATGAGACTGTTTGGCAGCGCCAGCAGC 600
 QY 791 ACCGCGCTTACCCAGCGGATCTTGATCCGCTGGGTGAGAGAGGCTAACAGAACAGA 850
 Db 601 ACCGCGCTTACCCAGCGGATCTTGATCCGCTGGGTGAGAGAGGCTAACAGAACAGA 660
 QY 851 GCGCGTGGCAGAGGAGATGCAAGGCTTGGGGGCAAAAGGAGACCTGTGAGACCAACCC 910
 Db 661 GCGCGTGGCAGAGGAGATGCAAGGCTTGGGGGCAAAAGGAGACCTGTGAGACCAACCC 720
 QY 911 TCAAGGGGACAGACCCCACTCACAACAGAGAGAGAGCAATATCAGACCCATGAC 970
 Db 721 TCAAGGGGACAGACCCCACTCACAACAGAGAGAGAGCAATATCAGACCCATGAC 780
 QY 971 CACACCCGCTTTCATCTGTATGATGCTGCTGTTTGGCTCCGATCTGAAGGCGCAGCTTC 1030
 Db 781 CACACCCGCTTTCATCTGTATGATGCTGCTGTTTGGCTCCGATCTGAAGGCGCAGCTTC 840
 QY 1031 GGGGCGCGCGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGG 1090
 Db 841 GGGGCGCGCGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGG 900
 QY 1091 CCACTTACCCCAAGGGGATGACCTGCGCCCGCCCAAGGAGAGGAGAGGAGAGGAGAGG 1150
 Db 901 CCACTTACCCCAAGGGGATGACCTGCGCCCGCCCAAGGAGAGGAGAGGAGAGGAGAGG 960
 QY 1151 CACCCAGCTGCTCCTCCAGAGACAGAGCGGGGCGAGCGGAGACCTCCAGAACTTATCT 1210
 Db 961 CACCCAGCTGCTCCTCCAGAGACAGAGCGGGGCGAGCGGAGACCTCCAGAACTTATCT 1020
 QY 1211 ATGGGTGGGTAACTCTTCAACGCGCCCTGAAAGGGGAGCTTCCCAATCCCTCAACCCAC 1270
 Db 1021 ATGGGTGGGTAACTCTTCAACGCGCCCTGAAAGGGGAGCTTCCCAATCCCTCAACCCAC 1080
 QY 1271 CTGATGTCCCAAGACACTGTATCAGACCAACAGTGCCCGCCCAAGAAATGAGGCTCAG 1330
 Db 1081 CTGATGTCCCAAGACACTGTATCAGACCAACAGTGCCCGCCCAAGAAATGAGGCTCAG 1140
 QY 1331 GATTTCAAGGCTTCAAGCTCAGGAGGTGACTTTCCGAGAGCCCTGTGTGATCTTCAAGGCT 1390
 Db 1141 GATTTCAAGGCTTCAAGCTCAGGAGGTGACTTTCCGAGAGCCCTGTGTGATCTTCAAGGCT 1200
 QY 1391 CGCTCAGTTAGCATTTAGTGGCATCTAACCCACAGAGAGGTGGGGGCAACCCAGAAACCA 1450
 Db 1201 CGCTCAGTTAGCATTTAGTGGCATCTAACCCACAGAGAGGTGGGGGCAACCCAGAAACCA 1260
 QY 1451 AAGCCCTTGAATGATATCTTTTATCAGGGTTCCTATGCGGCGCACCGGCGCAGAGT 1510

Db 1261 AAGCCCTTGAATGATATCTTTTATCAGGGTTCCTATGCGGCGCACCGGCGCAGAGT 1320
 QY 1511 ATGGCCCTTGGCAGAGGTAGAGACATTCATACCCAGGAAACCCAGATTAAGAA 1570
 Db 1321 ATGGCCCTTGGCAGAGGTAGAGACATTCATACCCAGGAAACCCAGATTAAGAA 1380
 QY 1571 GCGCCCTGAGGAGCAGACGATAGAGAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGG 1630
 Db 1381 GCGCCCTGAGGAGCAGACGATAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440
 QY 1631 TCTCTAGTCATTTCTTGCCTTTTCTCCGATTTGCGGATTTGGGGCGCACTCTAAGATG 1690
 Db 1441 TCTCTAGTCATTTCTTGCCTTTTCTCCGATTTGCGGATTTGGGGCGCACTCTAAGATG 1500
 QY 1691 CCTCTCTCCAGCCCTGTCTCAACCATCTCAATTAATAGTCCCAACCCAGGAGGCTGGCAC 1750
 Db 1501 CCTCTCTCCAGCCCTGTCTCAACCATCTCAATTAATAGTCCCAACCCAGGAGGCTGGCAC 1560
 QY 1751 CTCCCATCATCATTTGCTTGTCTGCAAGTGCAGATTAACCGGCGATTTGCCAACCTG 1810
 Db 1561 CTCCCATCATCATTTGCTTGTCTGCAAGTGCAGATTAACCGGCGATTTGCCAACCTG 1620
 QY 1811 G 1811
 Db 1621 G 1621

RESULT 7
 AK123762
 LOCUS
 DEFINITION
 ACCSSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS

1 Oka, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahara, K., Murakami, K., Yasuda, T., Iwayanagi, T., Waga, T., Shiratori, A., Sudo, H., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Niinomiya, K., Ishibashi, T., Yamauchi, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hirakawa, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yoshida, M., Houta, T., Kusano, J., Kanehori, K., Takahashi, Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Mutsaers, K., Yuuki, H., Oshima, A., Sasaki, N., Aotuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Saitoh, N., Sano, S., Moriya, S., Momiya, H., Sato, N., Takami, S., Teraoka, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, M., Watanabe, T., Sugiyama, Y., Takemoto, M., Kawakami, B., Yanahashi, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiya, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hiro, M., Ohmori, Y., Kawabata, A., Hiki, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsunaga, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shira, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuko, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isegaki, T., and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs. Nat. Genet. 36 (1), 40-45 (2004)

JOURNAL
 PUBMED

14702039

Db 1466 TATTAAAGAAAGCCCTCTGGGGGAGACAGACATAGAGGGGTGGAGTGGCTCCCTTT 1525
Qy 1621 ATCTGCAATCTCTAGTCATCTTCTGCTTTTCTCCCATTTGGGATTTGGGGCCAC 1680
Db 1526 ATCTGCAATCTCTAGTCATCTTCTGCTTTTCTCCCATTTGGGATTTGGGGCCAC 1585
Qy 1681 CTCGAATGCTCTCTCCAGCCCTGCTCAACCATCTCCAAATTAGTGCACACCCAGG 1740
Db 1586 CTCGAATGCTCTCTCCAGCCCTGCTCAACCATCTCCAAATTAGTGCACACCCAGG 1645
Qy 1741 GGCCTGACACCTCCACATGATCATTTGCTGTGCAAGTGCAGATAAAGCGCTGAT 1800
Db 1646 GGCCTGACACCTCCACATGATCATTTGCTGTGCAAGTGCAGATAAAGCGTGTGAT 1705
Qy 1801 TGCC 1804
Db 1706 TGCC 1709
RESULT 8
LOCUS BD158376 1614 bp DNA linear PAT 17-JAN-2003
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD158376
VERSION BD158376.1 GI:27864134
KEYWORDS JP 2002191363-A/13219.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1614)
AUTHORS Oca,T., Isogai,T., Nishikawa,T., Hayashi,K., Saio,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primer for synthesizing full-length cDNA and use thereof.
JOURNAL Patent: JP 2002191363-A 13219 09-JUL-2002;
HELIX RESEARCH INSTITUTE
COMMENT OS Homo sapiens (human)
PN JP 2002191363-A/13219
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
PI SAITO,
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI KEIICHI NAGAI,TETSUO OHSUKI
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
10, C12P21/02,C12O1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
FEATURES
Source Location/Qualifiers
1..1614
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 79.2%; Score 1434.8; DB 6; Length 1614;
Best Local Similarity 90.6%; Pred. No. 0;
Matches 1612; Conservative 0; Mismatches 2; Indels 166; Gaps 1;
Qy 25 AGTACACGACCTGAGTAGTAGGACACAGAGACCTCTCCATGTTTAGGACCTCT 84
Db 1 AGTACACGACCTGAGTAGTAGGACACAGAGACCTCTCCATGTTTAGGACCTCTCT 60
Qy 85 GGGGCTTACAGAGCTGGGCGCCGCTTGGGCGGACCTCCCATCTCGGCGGCGCGGATG 144
Db 61 GGGGCTTACAGAGCTGGGCGCCGCTTGGGCGGACCTCCCATCTCGGCGGCGCGGATG 120
Qy 145 GTCCGGGTCCGCTCCGAGTGTCTGTGCTCCCTGTTGTGCGGATGAGTGTGCG 204
|||||

Db 121 GTCCGGGCGCGCTCCGACAGTGTGTGCTGTCTCCCTGCTGTGGGTGCAAGTGTG 180
Qy 205 GTTCTGGGTTTCTGAGTTGGCGGGCCGTTTCAACGTAAGCTGTGGGCTCTCCGAGTGA 264
Db 181 GTTCTGGGTTTCTGAGTTGGCGGGCCGTTTCAACGTAAGCTGTGGGCTCTCCGAGTGA 240
Qy 265 GTCCGTCGCGCGGCTGCGCCCGGAGCGGCTTGTGCTGCGGGGGTCCGGGCGCCAGGC 324
Db 241 GTCCGTCGCGCGGCTGCGCCCGGAGCGGCTTGTGCTGCGGGGGTCCGGGCGCCAGGC 300
Qy 325 ATTCCGGGCTGACGATTGACGGGGAATCCGGAATGACACCGCGCCCTCCGCTTACCG 384
Db 301 ATTCCGGGCTGACGATTGACGGGGAATCCGGAATGACACCGCGCCCTCCGCTTACCG 360
Qy 385 ACGGTCGACGCTTGTGGGGAAGAAGTGGGGGAAGGGTCCCTGAGGATCCGATGCTT 444
Db 361 ACGGTCGACGCTTGTGGGGAAGAAGTGGGGGAAGGGTCCCTGAGGATCCGATGCTT 420
Qy 445 ACGAGCCAGATGCTCAGCTTTTATAGGTGACCTTACACATGTGACTTCACTCAGTTT 504
Db 421 ACGAGCCAGATGCTCAGCTTTTATAGGTGACCTTACACATGTGACTTCACTCAGTTT 480
Qy 505 GTGATCCGTAATAATGACAAATTCGAAGCTTCTTCAAGTCTGTGTGAGAGATTAAATG 564
Db 481 GTGATCCGTAATAATGACAAATTCGAAGCTTCTTCAAGTCTGTGTGAGAGATTAAATG 540
Qy 565 AATCAATGCTGTGTAAGCTCTTTTGCAGAGAGGAGCTCCGGAACAGGCGCTGGCGGCGAG 624
Db 541 AATCAATGCTGTGTAAGCTCTTTTGCAGAGAGGAGCTCCGGAACAGGCGCTGGCGGCGAG 600
Qy 625 AGCAGACCTGCTGTACACAGGAGACACAGGACATGAAAGCCCGCTGAGAGTGGCCGT 684
Db 601 AGCAGACCTGCTGTACACAGGAGACACAGGACATGAAAGCCCGCTGAGAGTGGCCGT 660
Qy 685 CAGTGGAGTACAGACCTTGGGCTTTCAGACACGCTGCCAGGTGCTTACCGGCTCAAAGC 744
Db 661 CAGTGGAGTACAGACCTTGGGCTTTCAGACACGCTGCCAGGTGCTTACCGGCTCAAAGC 720
Qy 745 CAGGAGCTCATATGATGATGAGACTCTGTTTGGACCCGACGAGGACCCGCGCTACCCC 804
Db 721 CAGGAGCTCATATGATGATGAGACTCTGTTTGGACCCGACGAGGACCCGCGCTACCCC 780
Qy 805 ACCGACTTGTGATCCGCTGGGTGAGAAAGGTAAACAGAACGAGGCGCTGGGCAAGA 864
Db 781 ACCGACTTGTGATCCGCTGGGTGAGAAAGGTAAACAGAACGAGGCGCTGGGCAAGA 840
Qy 865 GGCATCGAAGGCTTGGGGGCAAAAGGAGCTGTGAGACCACTCTCAAGGGGCAAGC 924
Db 841 GGCATCGAAGGCTTGGGGGCAAAAGGAGCTGTGAGACCACTCTCAAGGGGCAAGC 900
Qy 925 CCCCACCTTCAACCAAGAAAGAAAGAAATTCAGACCATGACGACACCCCGTCTTA 984
Db 901 CCCCACCTTCAACCAAGAAAGAAAGAAATTCAGACCATGACGACACCCCGTCTTA 960
Qy 985 CTGTGATGATGCTCTGTTTGGCTCCGATCTGAAGGCGGCAAGTCTGGGGCCCGCGGAT 1044
Db 961 CTGTGATGATGCTCTGTTTGGCTCCGATCTGAAGGCGGCAAGTCTGGGGCCCGCGGAT 1020
Qy 1105 GGGTATGCACTGGCCCGCCCGGAGGAGGACCACTGCGAGCATCCCACTGCTGTC 1164
Db 1081 GGGTATGCACTGGCCCGCCCGGAGGAGG----- 1109
Qy 1165 CTCGAAGACAGACCGGGGCGACGGGACACTCCAGAAATTATCTATGGGTGGATTACA 1224
Db 1110 ----- 1109
Qy 1225 CTCTTACGCCCCCTGAAGCGGAGCTTTCCATTCCCTCACCCACCTGAATGTCCAG 1284
Db 1110 ----- 1109

QY 1285 CACTGTCATCCAGCCACCAAGTGGCCCAACAATGGGCTTCAGATCTCAGGCTTC 1344
DB 1110 -----CACCAAGTGGCCCAACAATGGGCTTCAGATCTCAGGCTTC 1154
QY 1345 CACGTCAGGAGGTCACCTTCGGAGGCCCTTGTGACTTCAGAGGCTTCAGTACAT 1404
DB 1155 CACGTCAGGAGGTCACCTTCGGAGGCCCTTGTGACTTCAGAGGCTTCAGTACAT 1214
QY 1405 TTCAGTGCATCTACCCCAAGAGTGGGGCCACCCGAAACCAAGCCCTTGGAA 1464
DB 1215 TTCAGTGCATCTACCCCAAGAGTGGGGCCACCCGAAACCAAGCCCTTGGAA 1274
QY 1465 ATGATACCTTTCATCAGGGGTCCTATGGGGCCACGAGCAAGGATGGCCCTTGCA 1524
DB 1275 ATGATACCTTTCATCAGGGGTCCTATGGGGCCACGAGCAAGGATGGCCCTTGCA 1334
QY 1525 GGGTAGAGAGACATTCATCACCAGGAAACCCAGGATTTAAAGAACCCCTGTGGGGC 1584
DB 1335 GGGTAGAGAGACATTCATCACCAGGAAACCCAGGATTTAAAGAACCCCTGTGGGGC 1394
QY 1585 AGACAGATAGAGAGGAGGTCAGTGCCTCCCTTATTCCTGACATCTCTAGTGAATTC 1644
DB 1395 AGACAGATAGAGAGGAGGTCAGTGCCTCCCTTATTCCTGACATCTCTAGTGAATTC 1454
QY 1645 TTGCTTTTTCCTCCGATTTGGAGGAGCACTCTAGATGCTCTCCAGGCC 1704
DB 1455 TTGCTTTTTCCTCCGATTTGGAGGAGCACTCTAGATGCTCTCTCCAGGCC 1514
QY 1705 TGTCTCAACCATCTCCAAATTAAGTCCACCCAGAGGCTTGCACTTCCCATCATTC 1764
DB 1515 TGTCTCAACCATCTCCAAATTAAGTCCACCCAGAGGCTTGCACTTCCCATCATTC 1574
QY 1765 ATTGCTTGTGCGCAAGTCCGAATTAACGGCTGATTCG 1804
DB 1575 ATTGCTTGTGCGCAAGTCCGAATTAACGGCTGATTCG 1614

RESULT 9
AX880429
LOCUS AX880429 1614 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 15334 from Patent EP1074617.
ACCESSION AX880429
VERSION AX880429.1 GI:40035165
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS Ota,T., Iwagaki,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITL Prlments for synthesizing full-length cDNA and their use
JOURNML Patent: EP 1074617-A 15334 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
source
1. 1614
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
635..1168
/note="unamed protein product"

CDS

/codon_start=1
/protein_id="CAB90944.1"
/db_xref="GI:40035166"
/translation="MKTPELVAVSGMOTLGHQHRGSGYRVKARTSYVDETLFGSPAG
TRPTPPDFPWEKANTRGVAGKSGFALAKGSCETTPRGSTPTLTPKKQKVRP
ISHTPSYCDSEILFGSRSGAFSGAPRMAKGAALRALMLWTPPTPRSHSPRRREAP
VPTQMKRISGLRQRC"

ORIGIN

Query Match 79.2%; Score 1434.8; DB 6; Length 1614;
Best Local Similarity 90.6%; Pred. No. 0;

Matches 1612; Conservative 0; Mismatches 2; Indels 166; Gaps 1;

QY 25 AGTACACACGACCTGAGTGTGGACACAGAGGACCCCTTCATGTTAGGACCTCCT 84
DB 1 AGTACACACGACCTGAGTGTGGACACAGAGGACCCCTTCATGTTAGGACCTCCT 60
QY 85 GGGCTCAGAGAGCTGGAGCCCGCCCTTGGGCGGACTCCCATCCGCGGCGGAGATG 144
DB 61 GGGCTCAGAGAGCTGGAGCCCGCCCTTGGGCGGACTCCCATCCGCGGCGGAGATG 120
QY 145 GTCCGGTGCCTCCGACAGTGTCTGCTGCTCTCCCTGATGTCTGGTGCAGAAAGTCTGG 204
DB 121 GTCCGGTGCCTCCGACAGTGTCTGCTGCTCTCCCTGATGTCTGGTGCAGAAAGTCTGG 180
QY 205 GTTCTGGGTTTCTGATTCGCGGGCCGTTCAACATGACCTGTGCGGCTCTCCGAGTGA 264
DB 181 GTTCTGGGTTTCTGATTCGCGGGCCGTTCAACATGACCTGTGCGGCTCTCCGAGTGA 240
QY 265 GTCCGTCCGCGCGAGTCCCGGAGACGAGCTAGAGCTCCGAGGAGTCCGAGGACCCAGGC 324
DB 241 GTCCGTCCGCGCGAGTCCCGGAGACGAGCTAGAGCTCCGAGGAGTCCGAGGACCCAGGC 300
QY 325 ATTCCGGCTGCAATGACCGGAGATCCCGATGACCGCGGCCCGCGCTCAGCG 384
DB 301 ATTCCGGCTGCAATGACCGGAGATCCCGATGACCGCGGCCCGCGCTCAGCG 360
QY 385 ACGGTCACAGACTGTGTGGAGAAAGGTGCGGAGACGGGTCTCTAGAGATCCCATGCTT 444
DB 361 ACGGTCACAGACTGTGTGGAGAAAGGTGCGGAGACGGGTCTCTAGAGATCCCATGCTT 420
QY 445 ACGAGCCAAAGATGCTCAGCTTATATGATGTGACCTACATGTGACTTCACTCAGTTT 504
DB 421 ACGAGCCAAAGATGCTCAGCTTATATGATGTGACCTACATGTGACTTCACTCAGTTT 480
QY 505 GTGATCGTAAATGACAAATTCGAAGCTACTTCAAGTGTCTTGAAGAGATTTAAATG 564
DB 481 GTGATCGTAAATGACAAATTCGAAGCTACTTCAAGTGTCTTGAAGAGATTTAAATG 540
QY 565 AAACAAATGCTTGTAAAGCTCTTTTGGAGAGGAGCTTGGAAAGCAGGGCTTGGCCGAG 624
DB 541 AAACAAATGCTTGTAAAGCTCTTTTGGAGAGGAGCTTGGAAAGCAGGGCTTGGCCGAG 600
QY 625 AGCACACCTGTGTACACAGAGGACCAACAGGACGATGAAGACCCCGTGGAGCTGGCGT 684
DB 601 AGCACACCTGTGTACACAGAGGACCAACAGGACGATGAAGACCCCGTGGAGCTGGCGT 660
QY 685 CAGTGGATGACAGACCTTCGCTTCAAGACCGCTGCGAGAGTGGCTACCGGCTCAAGGC 744
DB 661 CAGTGGATGACAGACCTTCGCTTCAAGACCGCTGCGAGAGTGGCTACCGGCTCAAGGC 720
QY 745 CAGGACGTCAATGTGTGATGAGCTTGTGGACAGCCGACAGGACCCGCGCTTACCCC 804
DB 721 CAGGACGTCAATGTGTGATGAGCTTGTGGACAGCCGACAGGACCCGCGCTTACCCC 780
QY 805 ACCGACCTTCGATCGCGCTGGGTGGAGAGGCTTAACGAACCAAGAGCGTGGGCAAGGA 864
DB 781 ACCGACCTTCGATCGCGCTGGGTGGAGAGGCTTAACGAACCAAGAGCGTGGGCAAGGA 840
QY 865 GGCATCGAAGGCTTGGGGGCAAGAGGAGCTGTGAGACCAACCCCTCAAGGGGAGAGC 924
DB 841 GGCATCGAAGGCTTGGGGGCAAGAGGAGCTGTGAGACCAACCCCTCAAGGGGAGAGC 900
QY 925 CCCCACCTTCACACCAAGAAAGAAACAAATACAGACCCATACAGCACCCGCTTGA 984
DB 901 CCCCACCTTCACACCAAGAAAGAAACAAATACAGACCCATACAGCACCCGCTTGA 960
QY 985 CTGTGATGAGTGTGCTGTGTGCTCCCATCTGAAGAGGCGCAGCTTGGGGGCGCGGAT 1044
DB 961 CTGTGATGAGTGTGCTGTGTGCTCCCATCTGAAGAGGCGCAGCTTGGGGGCGCGGAT 1020
QY 1045 GGCAGAGGGGAGTCCGCAAGTCCGTGCTCTCTTGTGACGACCACTTACCCGAG 1104
DB 1021 GGCAGAGGGGAGTCCGCAAGTCCGTGCTCTCTTGTGACGACCACTTACCCGAG 1080


```

QY 1105 GGGTAGCAGCTCCGCCCCCGAGGGAGGACCACTGGAGCATTCACCCAGCTGGTCC 1164
DB 1081 GGGTAGCAGCTCCGCCCCCGAGGGAGG----- 1109
QY 1165 CTCGACAGACAGAGCCGGGGCAGCGGAGACTCCAGAAAGTTATCTATGGGTGGTACA 1224
DB 1110 ----- 1109
QY 1225 CTCTTCACGCCCCCTGAAAGGGGGAGCTTTTCCATTCCTCCACCACTGAATGTCCAG 1284
DB 1110 ----- 1109
QY 1285 CACTGTCATCCAGCCAGCAGTGGCCCCCACAACAAATGGGCTCAGGATCTCAGGCTTC 1344
DB 1110 -----CAGCAGTGGCCCCCACAACAAATGGGCTCAGGATCTCAGGCTTC 1154
QY 1345 CAGGTCAAGGGGTGACTTCCGAGAGCCCTGTGTGACTTCCAGGCTCGCTCAGTTAGCAT 1404
DB 1155 CAGGTCAAGGGGTGACTTCCGAGAGCCCTGTGTGACTTCCAGGCTCGCTCAGTTAGCAT 1214
QY 1405 TTCACTGTCATCTTACCCCAAGAGAGGTGGGGCCACCCAGAACCAAGCCCCCTTGGAA 1464
DB 1215 TTCACTGTCATCTTACCCCAAGAGAGGTGGGGCCACCCAGAACCAAGCCCCCTTGGAA 1274
QY 1465 ATGATACCTCTTATCAGGGTGGCTATGGGGCCAGGGCCAGAGTATGGGCCCTTGGCA 1524
DB 1275 ATGATACCTCTTATCAGGGTGGCTATGGGGCCAGGGCCAGAGTATGGGCCCTTGGCA 1334
QY 1525 GGGTAGAGGAGCATTCATCAACCCAGGAAACCCAGATTAAGAGACCCCTGTGGGGGC 1584
DB 1335 GGGTAGAGGAGCATTCATCAACCCAGGAAACCCAGATTAAGAGACCCCTGTGGGGGC 1394
QY 1585 AGACAGACATRGCAGGGGTGGGGCTGCTCCCTTTATCTCGAACAATCTCAGTGCATTC 1644
DB 1395 AGACAGACATRGCAGGGGTGGGGCTGCTCCCTTTATCTCGAACAATCTCAGTGCATTC 1454
QY 1645 TTGGCTTTTCTCCCGATTGGCGATTGGGGCCACCTCTAAGATGCTCTCTCCAGCCC 1704
DB 1455 TTGGCTTTTCTCCCGATTGGCGATTGGGGCCACCTCTAAGATGCTCTCTCCAGCCC 1514
QY 1705 TGTCTCAACCACTCTCCAAATTAATGTCGCAACCCAGGGGCTGTGGACCTCCCACTCATTC 1764
DB 1515 TGTCTCAACCACTCTCCAAATTAATGTCGCAACCCAGGGGCTGTGGACCTCCCACTCATTC 1574
QY 1765 ATTGCTCTTGGCCCAAGTGGGAAATAAACGGGTGATTTGCC 1804
DB 1575 ATTGCTCTTGGCCCAAGTGGGAAATAAACGGGTGATTTGCC 1614

RESULT 10
AK027741 1614 bp mRNA linear PRI 30-JAN-2004
LOCUS DEFINITION Homo sapiens cDNA FLJ14835 fis, clone OVARC1001369.
ACCESSION AK027741
VERSION AK027741.1 GI:14042642
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
Sekine, M., Odayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
Nagahari, A., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuna, M.,
Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,
Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,
Tanihara, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,
Miyakawa, K., Fujimori, K., Tanai, H., Kinata, M., Watanabe, M.,
Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,

```

```

Yosida, M., Horuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A.,
Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,
Takahashi, K., Arita, M., Imose, N., Mutsaers, K., Yuki, H., Oshima, A.,
Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,
Shiohata, N., Sano, S., Mori, S., Nomiyama, H., Satoh, N., Takami, S.,
Terahama, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,
Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,
Ozaki, K., Hirao, M., Omori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,
Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,
Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,
Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,
Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,
Nagase, T., Nomura, N., Kikuchi, H., Masuno, Y., Yamashita, R.,
Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T., and Sugano, S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
14702039

JOURNAL
PUBMED
REFERENCE
AUTHORS
TITILE
JOURNAL
AUTHORS
TITILE
COMMENT
FEATURES
SOURCE
CDS
ORIGIN
misc_feature
note="116 bases segment is present in AY007158."
Query Match 79.2%; Score 1434.8; DB 9; Length 1614;
Best Local Similarity 90.6%; Pred. No. 0;
Matches 1612; Conservative 0; Mismatches 2; Indels 166; Gaps 1;
25 AGTACACGACGACCTGAGTAGAGGAGGACGACGAGGACCTTCATGTTTGGGACCTCCT 84

```

Db 1 AGTACACAGCAGCTGATGATGAGCAACAGAGACCTCTTCATGTTTAAAGGACCTCTCT 60
 Qy 85 GGGGCTCAGAGAGCGTGGCGCGCCCTGAGGCGAGTCTCCGCCATCCGCGGGCGCAATG 144
 Db 61 GGGGCTCAGAGAGCGTGGCGCGCCCTGAGGCGAGTCTCCGCCATCCGCGGGCGCAATG 120
 Qy 145 GTCCGGGTCGCGTGGCGAGTGTGCTGCTGCTCCCTGAGTTCCTGAGGCAAAAGTCTGG 204
 Db 121 GTCCGGGTCGCGTGGCGAGTGTGCTGCTGCTCCCTGAGTTCCTGAGGCAAAAGTCTGG 180
 Qy 205 GTTCTGGGTTTGTGATTCGCGGGCCGTTTCAACAGTACGCTGTGCGGCTCTCTCGGTTA 264
 Db 181 GTTCTGGGTTTGTGATTCGCGGGCCGTTTCAACAGTACGCTGTGCGGCTCTCTCGGTTA 240
 Qy 265 GTCCGTCGCGCGCGGATGCGCGGAGCTAGGCTGCGCGGGAGTCCGCGGGCCCGAGGC 324
 Db 241 GTCCGTCGCGCGCGGATGCGCGGAGCTAGGCTGCGCGGGAGTCCGCGGGCCCGAGGC 300
 Qy 325 ATTCCGGGCTGCGATTTGACGCGGAGTCCCGAGTGCACGCGCGCGCCCGCTTCACCG 384
 Db 301 ATTCCGGGCTGCGATTTGACGCGGAGTCCCGAGTGCACGCGCGCGCCCGCTTCACCG 360
 Qy 385 AGGGTCCAGACCTGTGTGAGGAAAGAGTGTGCGGAGCGGCTCCCTGAGATCCCGATGCT 444
 Db 361 AGGGTCCAGACCTGTGTGAGGAAAGAGTGTGCGGAGCGGCTCCCTGAGATCCCGATGCT 420
 Qy 445 AGGAGCCAGATGCTCAGCTTATAGGTGTGACCTACACATGTAAGTCACTCACTCACTGTTT 504
 Db 421 AGGAGCCAGATGCTCAGCTTATAGGTGTGACCTACACATGTAAGTCACTCACTCACTGTTT 480
 Qy 505 GTGATCCGTAATAATGAGCAAAATTCAGAGTCACTTCAAGTGTGTTGAGAGATTTAAATG 564
 Db 481 GTGATCCGTAATAATGAGCAAAATTCAGAGTCACTTCAAGTGTGTTGAGAGATTTAAATG 540
 Qy 565 AAACATATGCTGTAAAGCTCTTTGCAAGAGGAGCTCTGGAAGACAGGCTGTGCGGCGAG 624
 Db 541 AAACATATGCTGTAAAGCTCTTTGCAAGAGGAGCTCTGGAAGACAGGCTGTGCGGCGAG 600
 Qy 625 AGCACACCTGCTGATCCAGAGGAGCAACAGGAGCAAGTAAAGCCCTGAGGAGCTGAGCTT 684
 Db 601 AGCACACCTGCTGATCCAGAGGAGCAACAGGAGCAAGTAAAGCCCTGAGGAGCTGAGCTT 660
 Qy 685 CAGTGGATGAGAGACCTCTGCGCTTCAAGACCGCTGCGAGGAGTCAACGCGGTCAAGGC 744
 Db 661 CAGTGGATGAGAGACCTCTGCGCTTCAAGACCGCTGCGAGGAGTCAACGCGGTCAAGGC 720
 Qy 745 CAGGAGTCAATGATGATGAGATGCTGTTTGGACGCCAGGAGCAACCGGCTTACCC 804
 Db 721 CAGGAGTCAATGATGATGAGATGCTGTTTGGACGCCAGGAGCAACCGGCTTACCC 780
 Qy 805 ACCGAGCTTTCATCCGCTGCTGAGTGAAGGCTTAAACAGAACAGGCTGTGGGCAAGA 864
 Db 781 ACCGAGCTTTCATCCGCTGCTGAGTGAAGGCTTAAACAGAACAGGCTGTGGGCAAGA 840
 Qy 865 GGCATCGAAGGCTTGGGGGCAAAAGGAGCTGTGAGACCAACCCCTTCAAGGGAGAGCAC 924
 Db 841 GGCATCGAAGGCTTGGGGGCAAAAGGAGCTGTGAGACCAACCCCTTCAAGGGAGAGCAC 900
 Qy 925 CCCCACTCTCAACCAAGAGAGAAACAATAATACAGACCCATCAGCCACCCCGCTCTTA 984
 Db 901 CCCCACTCTCAACCAAGAGAGAAACAATAATACAGACCCATCAGCCACCCCGCTCTTA 960
 Qy 985 CTGAGATGAGTCCGCTGTTTGGTCCCGATCTGAAGGCGCAGCTTGGGGGCGCCGCGAT 1044
 Db 961 CTGAGATGAGTCCGCTGTTTGGTCCCGATCTGAAGGCGCAGCTTGGGGGCGCCGCGAT 1020
 Qy 1045 GGCAGAGGGAGATGCGCAAAAGCTCCGATCTCTTGTGAGAGGCAACCACTAACCCCGAG 1104
 Db 1021 GGCAGAGGGAGATGCGCAAAAGCTCCGATCTCTTGTGAGAGGCAACCACTAACCCCGAG 1080
 Qy 1105 GGGTAGCAGCTTGGCGCGCGCCCGAGGAGGACCACTGTGAGGCACTTCAACCACTGTGTC 1164
 Db 1081 GGGTAGCAGCTTGGCGCGCGCCCGAGGAGG----- 1109

Qy 1165 CTCCAAGACAGAGCGGGGCGAGCGGACAGACTCCAGAAATATATGATGATGAGTTACA 1224
 Db 1110 ----- 1109
 Qy 1225 CTCTTCAGCGCCCTGAAAGCGGGGACTTTCCTATTCCTCACCACCTGAATGTCCGAG 1284
 Db 1110 ----- 1109
 Qy 1285 CACTGATATCCAGCCACCAAGTGTGCCCCCACAACAATGGGCTCAGATCTCAGGCTTC 1344
 Db 1110 -----CACCAAGTGTGCCCCCACAACAATGGGCTCAGATCTCAGGCTTC 1154
 Qy 1345 CACGTCAAGGGTGAACCTTCCGAGGCGCCCTGATGACTTTCAGAGGCTGCTCAGTATGAT 1404
 Db 1155 CACGTCAAGGGTGAACCTTCCGAGGCGCCCTGATGACTTTCAGAGGCTGCTCAGTATGAT 1214
 Qy 1405 TTCACTGCTCATTCACCCACGACGAGTGGGGCAACCCAGAAACCAAGGCCCTTGGAA 1464
 Db 1215 TTCACTGCTCATTCACCCACGACGAGTGGGGCAACCCAGAAACCAAGGCCCTTGGAA 1274
 Qy 1465 ATGATATCTTTCATCAGGCTTGTCTATGAGGCGCACGCGGACAGATGTGCCCCCTTGCA 1524
 Db 1275 ATGATATCTTTCATCAGGCTTGTCTATGAGGCGCACGCGGACAGATGTGCCCCCTTGCA 1334
 Qy 1525 GGGTAGAGGACATTCATCACCCAGGAAACCCAGATTTAAAGAGGCCCTGTGGGGG 1584
 Db 1335 GGGTAGAGGACATTCATCACCCAGGAAACCCAGATTTAAAGAGGCCCTGTGGGGG 1394
 Qy 1585 AGACAGATATGACAGAGGAGTGGGCACTGCTCCCTTATCTGACATCTTATGATGATTC 1644
 Db 1395 AGACAGATATGACAGAGGAGTGGGCACTGCTCCCTTATCTGACATCTTATGATGATTC 1454
 Qy 1645 TTGCTTTTCTTCCGATTTGGGATTTGGGGGCACTCTTAAAGATGCTCTTCCAGGCC 1704
 Db 1455 TTGCTTTTCTTCCGATTTGGGATTTGGGGGCACTCTTAAAGATGCTCTTCCAGGCC 1514
 Qy 1705 TGTCTCAACCAATCTCCCAATTTAGTGGCAACCCAGGGGCTGGGACCTCCCAATCATTC 1764
 Db 1515 TGTCTCAACCAATCTCCCAATTTAGTGGCAACCCAGGGGCTGGGACCTCCCAATCATTC 1574
 Qy 1765 ATTGTCTTGTCCCAAGTGGCAATTAACGGGCTGATGCC 1804
 Db 1575 ATTGTCTTGTCCCAAGTGGCAATTAACGGGCTGATGCC 1614
 RESULT 11
 BC022092 1407 bp mRNA linear PRI 29-JUN-2004
 LOCUS BC022092
 DEFINITION Homo sapiens hypothetical protein FLJ14827, mRNA (cDNA clone
 MGC:4570 IMAGE:3029785), complete cds.
 ACCESSION BC022092
 VERSION BC022092.2 GI:33870282
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1407)
 REFERENCES
 1. Aleschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Wax, S.I., Wang, J., Hsieh, F.,
 Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Udell, T.B., Toshitaki, S.,
 Carninci, P., Prange, C., Raha, S., Loughran, N.A., Peters, G.J.,
 Abramson, R.D., Mullen, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worsley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huij, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywicki, M.I., Skalska, U., Snailus, D.E.,
 Schmechel, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 1407)
 Strausberg, R.
 Direct Submission
 Submitted (22-JUN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 On Aug 19, 2003 this sequence version replaced gi:18314382.
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: DCTD/DTP
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadansystemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Ketterman, Anuradha
 Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LIML at: <http://image.liml.gov>
 Series: IRAL Plate: 10 Row: h Column: 14
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 14249569.
 Location/Qualifiers
 1. 1407
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:4570 IMAGE:3029785"
 /tissue_type="Ovary, adenocarcinoma"
 /clone_id="NIH_MGC_9"
 /lab_host="DH10B-R"
 /note="Vector: POTB7"
 1. 1407
 /gene="FLJ14827"
 /db_xref="LocusID:84934"
 224. 1033
 /gene="FLJ14827"
 /codon_start=1
 /product="hypothetical protein FLJ14827"
 /protein_id="AAH22092.1"
 /db_xref="GI:18314383"
 /db_xref="LocusID:84934"
 /translation="MKTPELVASGMOTLQIQRRCRGYVKAQTSTVYDETLFGSPAG
 IRLTPPDPFPWVEKARTRGVGKESKALGAKGSCETPSRSTPLTPRKKNKYP
 ISHPSPYCDLSFGSRSESGAFAPAKMAALRALMTPTPTPTGSSPBRBP
 LRTHPAGPSTKTERPAADSOIKSMGSLSRPLKSLSTLIHNPSTGHATISAP
 HTNPQDLRSTSGVTFRSPLVTSRAVSISVPSITRRGATQKRPWK"
 ORIGIN
 Query Match 75.7%; Score 1371; DB 9; Length 1407;
 Best Local Similarity 100.0%; Pred. No. 5,1e-300;
 Matches 1371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 121 GATTAATGAAACAATGCTTTGTAAGCTTTTACAGAGGAGCCTCGAAGCAGGCGCT
 Qy 616 GGGCGGAGAGCACCCTGTGTACACAGGAGCACAAGCAGATGAAGCCCCCGTGA
 Db 181 GGGCGGAGAGCACCCTGTGTACACAGGAGCACAAGCAGATGAAGCCCCCGTGA
 Qy 676 GCTGCGGTGAGGAGATGAGACCTTGGCCTTCAGACCGGTGCGAGGTGACTACG
 Db 241 GCTGCGGTGAGGAGATGAGACCTTGGCCTTCAGACCGGTGCGAGGTGACTACG
 Qy 736 GGTCAAGGCGAGACGTCAATATGATGATGAGACTTGTGTCAGCCGACAGGACCG
 Db 301 GGTCAAGGCGAGACGTCAATATGATGATGAGACTTGTGTCAGCCGACAGGACCG
 Qy 796 GCTTACCCCAACCGGACTTGCATCCGCGGTGAGGAGGCTTAACAGAACAGGCGT
 Db 361 GCTTACCCCAACCGGACTTGCATCCGCGGTGAGGAGGCTTAACAGAACAGGCGT
 Qy 856 GGGCAAGAGAGCATGCAAGGCTTGGGGGCAAGGGAGCTGTAGACACCCCTCAAG
 Db 421 GGGCAAGAGAGCATGCAAGGCTTGGGGGCAAGGGAGCTGTAGACACCCCTCAAG
 Qy 916 GGGCAGACCCCAACCTCAACCAAGAAAGAAACAATTAAGACCCATCAAGCAGC
 Db 481 GGGCAGACCCCAACCTCAACCAAGAAAGAAACAATTAAGACCCATCAAGCAGC
 Qy 976 CCGGTCTTACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG
 Db 541 CCGGTCTTACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG
 Qy 1036 CCGCGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
 Db 601 CCGCGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
 Qy 1096 TACCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
 Db 661 TACCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
 Qy 1156 AGCTGTGCTTCCAAACACAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
 Db 721 AGCTGTGCTTCCAAACACAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
 Qy 1216 TGGATTACACTTTCACGCCCCCTGAGAGCGGAGACTTTCATTCCTCACCACCTGA
 Db 781 TGGATTACACTTTCACGCCCCCTGAGAGCGGAGACTTTCATTCCTCACCACCTGA
 Qy 1276 TGTCCCAAGACATGTCATCAGCCCAAGTGCCTCCCAACAATGAGCTCAGATCT
 Db 841 TGTCCCAAGACATGTCATCAGCCCAAGTGCCTCCCAACAATGAGCTCAGATCT
 Qy 1336 CAGGCTTCCAGCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
 Db 901 CAGGCTTCCAGCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
 Qy 1396 AGTTAGACTTTCAGTGCATCTACCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG
 Db 961 AGTTAGACTTTCAGTGCATCTACCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG
 Qy 1456 CCCTTGAATGATTAATCTTTTCAAGGTTGCTTATGAGGAGCAGGAGCAGGATGAGC
 Db 1021 CCCTTGAATGATTAATCTTTTCAAGGTTGCTTATGAGGAGCAGGAGCAGGATGAGC
 Qy 1516 CCCTTGCAGAGGATGAGAGCATTCAATCAGGAGGAGGAGGAGGAGGAGGAGGAGG
 Db 1081 CCCTTGCAGAGGATGAGAGCATTCAATCAGGAGGAGGAGGAGGAGGAGGAGGAGG
 Qy 1576 TGTGAGGAGGAGAGCATTATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
 Db 1141 TGTGAGGAGGAGAGCATTATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
 Qy 1636 AGTGCATCTTGCCTTTTCTCCGATGTCGATTTGGGAGCAGCTTAAGATGCTCT

Db 1201 AGTCGATCTTCTGCTTTTCTCCCGATTGCGGATTTGGGGGCGCACCCTTAAGATGCTCT 1260

Qy 1696 CTCGAGCCCTGCTCTACACCATCTCAATTAATAGTGCACACGAGGGGCGCTGGACCTCC 1755

Db 1261 CTCGAGCCCTGCTCTACACCATCTCAATTAATAGTGCACACGAGGGGCGCTGGACCTCC 1320

Qy 1756 ACATCATCATCTGCTCTGCTGCACAGTGCAGATTAACGGCGTATGCCAA 1806

Db 1321 ACATCATCATCTGCTCTGCTGCACAGTGCAGATTAACGGCGTATGCCAA 1371

RESULT 12
AC089999/C
LOCUS
DEFINITION Homo sapiens 12 BAC RP11-545P7 (Roswell Park Cancer Institute Human BAC library) complete sequence.
AC089999 AC021161
VERSION AC089999.20 GI:21617628
KEYWORDS HTG.
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C., Albright,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.U., Ding,Y., Dinu,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Escotto,M., Falle,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Hollway,C., Hollins,B., Homsl,F., Howard,S., Huber,J., Huliy,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korrah,U., Kovar,C., Kravovic,J., Kurehli,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W., Loulesged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapa,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mashiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S., Metcalf,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Monabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,N., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S., Oguh,M., Okunnu,G., Oregunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,B., Pul,L., Quiles,M., Ren,Y., Rivers,M., Rojas,A., RojudoKan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shoshchail,N., Sisson,I., Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Thomas,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Umaní,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wlarczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R., Weinstock,G. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 199522)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

REFERENCE BAYLOR PLAZA, HOUSTON, TX 77030, USA
AUTHORS 3 (bases 1 to 199522)
TITLE Direct Submission
JOURNAL Submitted (21-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE BAYLOR PLAZA, HOUSTON, TX 77030, USA
AUTHORS 4 (bases 1 to 199522)
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE BAYLOR PLAZA, HOUSTON, TX 77030, USA
AUTHORS 5 (bases 1 to 199522)
TITLE Direct Submission
JOURNAL Submitted (28-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE BAYLOR PLAZA, HOUSTON, TX 77030, USA
AUTHORS 6 (bases 1 to 199522)
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jun 27, 2002 this sequence version replaced gi:21490131.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht>

FEATURES
source Location/Qualifiers
1..199522
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-545P7"
complement(25..342)
/rpt_family="AluSx"
complement(52..2005)
repeat_region
misc_feature

```

/note="overlaps bases 1..1954 of clone AC127004"
/function="clone overlap"
STS
repeat_region /strand.name="SHGC-148490"
complement(943..1150)
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
complement(1487..1598)
/rpt_family="MIR"
2041..2731
/rpt_family="L2"
complement(5385..5699)
/rpt_family="AluY"
complement(5720..6033)
/rpt_family="AluB"
5755..5967
STS
repeat_region /strand.name="SHGC-67307"
complement(7231..7531)
/rpt_family="AluSq"
complement(7633..7927)
/rpt_family="AluSx"
complement(7986..8413)
/rpt_family="L2"
complement(8408..8611)
/rpt_family="MIR"
complement(9109..9166)
/rpt_family="L2"
9207..9256
/rpt_family="(CA)n"
9566..10270
/rpt_family="AluSq"
complement(11877..12165)
/rpt_family="AluB"
12422..12500
/rpt_family="AluS"
13357..13658
/rpt_family="AluSx"
complement(14062..14304)
/rpt_family="MIR"
complement(14379..14482)
/rpt_family="AluSg/x"
14642..14738
/rpt_family="MER102"
15557..15798
STS
repeat_region /strand.name="Sts-AA001615"
17094..17319
/rpt_family="MIR"
complement(18276..18388)
/rpt_family="MIR"
19450..19813
/rpt_family="MLT1B"
20261..20366
/rpt_family="AluSx"
20367..20547
/rpt_family="MLT1B"
20548..20733
/rpt_family="AluSx"
complement(21056..21223)
/rpt_family="L2"
21224..21514
/rpt_family="AluSx"
complement(21515..21595)
/rpt_family="L2"
22244..22372
/rpt_family="MIR"
complement(22394..22690)
/rpt_family="AluB"
22708..22777
Query March 47.2% Score 854; DB 9; Length 199522;
Best Local Similarity 100.0%; Pred. No. 6.8e-183;
Matches 854; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
958 CAGACCATGACGACACCCGCTTACTGTGATGAGTCGCTGTTGGCTCCGATCTGA 1017

```

```

|||||
Db 59708 CAGACCATGACGACACCCGCTTACTGTGATGAGTCGCTGTTGGCTCCGATCTGA 59649
Qy 1018 AGGCGCAGCTTGGGGGCCCCGGGATGGGAGGGGGAGTGGCGAAAGCTCCGTGCTT 1077
Db 59648 AGGCGCAGCTTGGGGGCCCCGGGATGGGAGGGGGAGTGGCGAAAGCTCCGTGCTT 59589
Qy 1078 CTGTGGAGCCACACACCTTACCCCGAGGGGTAGCCACTCGCCCCGCCCGACGAGGACCC 1137
Db 59588 CTGTGGAGCCACACACCTTACCCCGAGGGGTAGCCACTCGCCCCGCCCGACGAGGACCC 59529
Qy 1138 ACTGCGAGCATTACACCTGCTGCTTCCCAAGACAGAGCCGGGGCAGCGGACATCTC 1197
Db 59528 ACTGCGAGCATTACACCTGCTGCTTCCCAAGACAGAGCCGGGGCAGCGGACATCTC 59469
Qy 1198 CCGAAGATTATCATGGGTGGGTATACCTTACAGGCCCCCTGAAGGGGGAGCTTCCCA 1257
Db 59468 CCGAAGATTATCATGGGTGGGTATACCTTACAGGCCCCCTGAAGGGGGAGCTTCCCA 59409
Qy 1258 TTCCCTCACCCACCTGATGTCCCGACAGCTGGTATCCAGCCACAGTGCCTCCACAC 1317
Db 59408 TTCCCTCACCCACCTGATGTCCCGACAGCTGGTATCCAGCCACAGTGCCTCCACAC 59349
Qy 1318 AAATGGGCTCAGAGATCTCAGGCTTCCAGCTCAGGGGTGAACCTTCCGAGCCCTGCT 1377
Db 59348 AAATGGGCTCAGAGATCTCAGGCTTCCAGCTCAGGGGTGAACCTTCCGAGCCCTGCT 59289
Qy 1378 GACTTCAGGGGCTCGCTCAGTATGACATTCAGTGCATTCACCCAGAGAGTGGGGG 1437
Db 59288 GACTTCAGGGGCTCGCTCAGTATGACATTCAGTGCATTCACCCAGAGAGTGGGGG 59229
Qy 1438 CACCCAGAAACCAAGGCCCTTGAATGATCTTTTCATCAGGGTTCCTATGGGGG 1497
Db 59228 CACCCAGAAACCAAGGCCCTTGAATGATCTTTTCATCAGGGTTCCTATGGGGG 59169
Qy 1498 CACGGGACAGTATGGCCCCCTTGGAGGTAGAGAGACATTCATCACCCAGGAAACCC 1557
Db 59168 CACGGGACAGTATGGCCCCCTTGGAGGTAGAGAGACATTCATCACCCAGGAAACCC 59109
Qy 1558 AGGTATTAAGAAAGCCCTGTGGGGGAGACATAGAGGGGGGAGTGGCTCCC 1617
Db 59108 AGGTATTAAGAAAGCCCTGTGGGGGAGACATAGAGGGGGGAGTGGCTCCC 59049
Qy 1618 TTATCTGACAAATCTAGTCGATTCCTGCTTTTTCCTCCGATTTGGGGG 1677
Db 59048 TTATCTGACAAATCTAGTCGATTCCTGCTTTTTCCTCCGATTTGGGGG 58989
Qy 1678 CACCTTAGATGCTCTCTCCAGCCCTGTCCACACCATTCGAAATTAATGTCACACC 1737
Db 58988 CACCTTAGATGCTCTCTCCAGCCCTGTCCACACCATTCGAAATTAATGTCACACC 58929
Qy 1738 AGGGGCTGGCACCTCCACATCATCATGCTGTGTCGCAAGTGGGAATAAGCGCGT 1797
Db 58928 AGGGGCTGGCACCTCCACATCATCATGCTGTGTCGCAAGTGGGAATAAGCGCGT 58869
Qy 1798 GATTGCCAACCTGG 1811
Db 58868 GATTGCCAACCTGG 58855

RESULT 13
BD149092 784 bp DNA linear PAT 17-JAN-2003
LOCUS BD149092
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD149092
VERSION BD149092.1 GI:27854850
KEYWORDS JP 2002191363-A/3935.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euleleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 784)
Oca.T., Isogai.T., Nishikawa.T., Hayaishi.K., Saito.K., Yamamoto.T.,
AUTHORS

```

TITLE
JOURNAL
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 3935 09-JUL-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002191363-A/3935

COMMENT
PD 09-JUL-2002
PI 28-JUL-2000 JP 2000280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
PI SAITO,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUJI OTSUKI
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10,
PC C12P21/02, C12O1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT source 1.784
Location/Qualifiers
FT source 1.784
/organism='Homo sapiens (human)'.
location/Qualifiers
1.784
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'

ORIGIN

Query Match 36.5%; Score 661.4; DB 6; Length 784;
Best Local Similarity 98.7%; Pred. No. 3.7e-139;
Matches 687; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

QY 191 GTGCAAGTGTGCTGGGTTCTGAGTTCTGAGTTCGCGGGCCGTTCAACAGTACCTGTGCC 250
Db 1 GTGCAAGTGTGCTGGGTTCTGAGTTCTGAGTTCGCGGGCCGTTCAACAGTACCTGTGCC 60
QY 251 GGTCTCTGGGATGATCGGTCCGCGCGCGGTCGCCCGGAGACGAGCTAGGCTGCGGGGAT 310
Db 61 GGTCTCTGGGATGATCGGTCCGCGCGCGGTCGCCCGGAGACGAGCTAGGCTGCGGGGAT 120
QY 311 CCGGGGCCCCAGGCAATTCGCGGCTGCAATTTGACGCGGATCCCGGATGACCGCGCGCC 370
Db 121 CCGGGGCCCCAGGCAATTCGCGGCTGCAATTTGACGCGGATCCCGGATGACCGCGCGCC 180
QY 121 CCGGGGCCCCAGGCAATTCGCGGCTGCAATTTGACGCGGATCCCGGATGACCGCGCGCC 180
Db 121 CCGGGGCCCCAGGCAATTCGCGGCTGCAATTTGACGCGGATCCCGGATGACCGCGCGCC 180
QY 371 CCGGCGCCTCAACGAGCGGATCCAGACCTGTGGGAAAGGTCGCGGAGCGGGATCCCTGA 430
Db 181 CCGGCGCCTCAACGAGCGGATCCAGACCTGTGGGAAAGGTCGCGGAGCGGGATCCCTGA 240
QY 431 GGATCCGATGCTTGAATCCGTAATGAGCAATTTGAAAGCTTACATGTGAC 490
Db 241 GGATCCGATGCTTGAATCCGTAATGAGCAATTTGAAAGCTTACATGTGAC 300
QY 491 TTCACCTGATTTGTGATCCGTAATGAGCAATTTGAAAGCTTACATGTGAC 550
Db 301 TTCACCTGATTTGTGATCCGTAATGAGCAATTTGAAAGCTTACATGTGAC 360
QY 551 GAGAGATTAATGAAACAATGCTTTGAAAGCTTTTCAAGAGGAGCCTCGAAGAG 610
Db 361 GAGAGATTAATGAAACAATGCTTTGAAAGCTTTTCAAGAGGAGCCTCGAAGAG 420
QY 611 GGCTGGCGCGAGACACCTGCTGACCAAGGAGCAACAGGAGATGAAAGACCCC 670
Db 421 GGCTGGCGCGAGACACCTGCTGACCAAGGAGCAACAGGAGATGAAAGACCCC 480
QY 671 GTGAGCTGGCGCTGATGGAGATGACACCTCGGCTTTCAGACCGCTGCGAGGTGC 730
Db 481 GTGAGCTGGCGCTGATGGAGATGACACCTCGGCTTTCAGACCGCTGCGAGGTGC 540
QY 731 TACCGGGTCAAGGCGAGGACGTCATATGTGATG-AGA-CTGTTTGGCAGCCGAGG 789
Db 541 TACCGGGTCAAGGCGAGGACGTCATATGTGATG-AGA-CTGTTTGGCAGCCGAGG 600
QY 790 CACCGGCTTACCCAGCGGATTCGATCCGCTGGGTGGAGAGGCTAACAGAACGAG 849

Db 601 CACCGGCTTACCCAGCGGATTCGATTCGCTGGGTGGAGAGGCTAACAGAACGAG 660
QY 850 AGGCGTGGGCAAGGA-GGCATGCAAGGCTTGGGG 884
Db 661 AGGCGTGGGCAAGGAGGATTCAGAGGCTTGGGG 696

RESULT 14
AX869030
LOCUS
DEFINITION Sequence 3935 from Patent EP1074617.
ACCESSION AX869030
VERSION AX869030.1 GI:40023893
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
Primer for synthesizing full-length cDNA and their use
Patent: EP 1074617-A 3935 07-FEB-2001;
Research Association for Biotechnology (JRP)
Location/Qualifiers
1.784
/organism='Homo sapiens'
/mol_type='unassigned DNA'
/db_xref='taxon:9606'

ORIGIN

Query Match 36.5%; Score 661.4; DB 6; Length 784;
Best Local Similarity 98.7%; Pred. No. 3.7e-139;
Matches 687; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

QY 191 GTGCAAGTGTGCTGGGTTCTGAGTTCTGAGTTCGCGGGCCGTTCAACAGTACCTGTGCC 250
Db 1 GTGCAAGTGTGCTGGGTTCTGAGTTCTGAGTTCGCGGGCCGTTCAACAGTACCTGTGCC 60
QY 251 GGTCTCTGGGATGATCGGTCCGCGCGCGGTCGCCCGGAGACGAGCTAGGCTGCGGGGAT 310
Db 61 GGTCTCTGGGATGATCGGTCCGCGCGCGGTCGCCCGGAGACGAGCTAGGCTGCGGGGAT 120
QY 311 CCGGGGCCCCAGGCAATTCGCGGCTGCAATTTGACGCGGATCCCGGATGACCGCGCGCC 370
Db 121 CCGGGGCCCCAGGCAATTCGCGGCTGCAATTTGACGCGGATCCCGGATGACCGCGCGCC 180
QY 121 CCGGGGCCCCAGGCAATTCGCGGCTGCAATTTGACGCGGATCCCGGATGACCGCGCGCC 180
Db 121 CCGGGGCCCCAGGCAATTCGCGGCTGCAATTTGACGCGGATCCCGGATGACCGCGCGCC 180
QY 371 CCGGCGCCTCAACGAGCGGATCCAGACCTGTGGGAAAGGTCGCGGAGCGGGATCCCTGA 430
Db 181 CCGGCGCCTCAACGAGCGGATCCAGACCTGTGGGAAAGGTCGCGGAGCGGGATCCCTGA 240
QY 431 GGATCCGATGCTTGAATCCGTAATGAGCAATTTGAAAGCTTACATGTGAC 490
Db 241 GGATCCGATGCTTGAATCCGTAATGAGCAATTTGAAAGCTTACATGTGAC 300
QY 491 TTCACCTGATTTGTGATCCGTAATGAGCAATTTGAAAGCTTACATGTGAC 550
Db 301 TTCACCTGATTTGTGATCCGTAATGAGCAATTTGAAAGCTTACATGTGAC 360
QY 551 GAGAGATTAATGAAACAATGCTTTGAAAGCTTTTCAAGAGGAGCCTCGAAGAG 610
Db 361 GAGAGATTAATGAAACAATGCTTTGAAAGCTTTTCAAGAGGAGCCTCGAAGAG 420
QY 611 GGCTGGCGCGAGACACCTGCTGACCAAGGAGCAACAGGAGATGAAAGACCCC 670
Db 421 GGCTGGCGCGAGACACCTGCTGACCAAGGAGCAACAGGAGATGAAAGACCCC 480
QY 671 GTGAGCTGGCGCTGATGGAGATGACACCTCGGCTTTCAGACCGCTGCGAGGTGC 730
Db 481 GTGAGCTGGCGCTGATGGAGATGACACCTCGGCTTTCAGACCGCTGCGAGGTGC 540
QY 731 TACCGGGTCAAGGCGAGGACGTCATATGTGATG-AGA-CTGTTTGGCAGCCGAGG 789

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1811; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AGTGCTGCGGCTCTCGGCGGCTTACACAGCCTGATGAGTGGCAACAAGAGAC 60
Db 1 AGTGCTGCGGCTCTCGGCGGCTTACACAGCCTGATGAGTGGCAACAAGAGAC 60
QY 61 CCTCTCATGTTTAGAGGACCTCCCTGAGGCTCAGAGAGCTGAGCGGCGGCTCTGAGGAGAC 120
Db 61 CCTCTCATGTTTAGAGGACCTCCCTGAGGCTCAGAGAGCTGAGCGGCGGCTCTGAGGAGAC 120
QY 121 TCCGCCCATCCGCGGCGGCAATGCTCGGAGTCCGCTCGCAGTGTCTGCTGCTCCC 180
Db 121 TCCGCCCATCCGCGGCGGCAATGCTCGGAGTCCGCTCGCAGTGTCTGCTGCTCCC 180
QY 181 TGGTGTGCTGGGTGCAAAAGTCTGGGTTCTGGGTTTCTGGAATGCGCGGCGGCTTCAACGT 240
Db 181 TGGTGTGCTGGGTGCAAAAGTCTGGGTTCTGGGTTTCTGGAATGCGCGGCGGCTTCAACGT 240
QY 241 AGCTGTGCGGCTCTCGGAGTGAAGTCCGCGCGCGGCTGCGCGGCGGCGGCTGAGG 300
Db 241 AGCTGTGCGGCTCTCGGAGTGAAGTCCGCGCGCGGCTGCGCGGCGGCGGCTGAGG 300
QY 301 TGCCTGCGGAGTCCGCGGCGGCAAGCATTCCGAGCTGCAATGACGCGGATCCCGATGCA 360
Db 301 TGCCTGCGGAGTCCGCGGCGGCAAGCATTCCGAGCTGCAATGACGCGGATCCCGATGCA 360
QY 361 CCGCGGCGGCGGCGGCTCTCAACGAGCGGCTCAGACCTGCTGGGAGAGAGAGTCCGCGGAC 420
Db 361 CCGCGGCGGCGGCGGCTCTCAACGAGCGGCTCAGACCTGCTGGGAGAGAGAGTCCGCGGAC 420
QY 421 GGGTCCCTGAGATCCCGATGCTCAGAGCGCAAGATCTCAGCTTTATAGTGTGACTA 480
Db 421 GGGTCCCTGAGATCCCGATGCTCAGAGCGCAAGATCTCAGCTTTATAGTGTGACTA 480
QY 481 CACATGTGACTTCACTCACTGATTTTGTGATCCGTAATAATGACAATAATGCAAGCTACTTCA 540
Db 481 CACATGTGACTTCACTCACTGATTTTGTGATCCGTAATAATGACAATAATGCAAGCTACTTCA 540
QY 541 CAGTGTCTGTGAAGATTAATGAACATATCTGTGAAGCTTTTGAAGCTTTTGAAGAGGAGGCC 600
Db 541 CAGTGTCTGTGAAGATTAATGAACATATCTGTGAAGCTTTTGAAGCTTTTGAAGAGGAGGCC 600
QY 601 TCGGAAGCAGAGGCTGCGGCGGCAAGACACTGCTGTCAACAAGGAGCAACAAGGACAT 660
Db 601 TCGGAAGCAGAGGCTGCGGCGGCAAGACACTGCTGTCAACAAGGAGCAACAAGGACAT 660
QY 661 GAAGACCCCGTGAAGTGGCCGTCAAGTGGATGACAGACCTTGGGCTTCAACAACGCTG 720
Db 661 GAAGACCCCGTGAAGTGGCCGTCAAGTGGATGACAGACCTTGGGCTTCAACAACGCTG 720
QY 721 CCGAGGTGCTACCGGGTCAAGGCGAGGATGATATGATGATGAGCTGTGTTGGCAG 780
Db 721 CCGAGGTGCTACCGGGTCAAGGCGAGGATGATATGATGATGAGCTGTGTTGGCAG 780
QY 781 CCGAGCAGAGCACCAGGCTTACCCACCGAGCTTGCATCCGCTCGGCTGAGGAGGCTTAA 840
Db 781 CCGAGCAGAGCACCAGGCTTACCCACCGAGCTTGCATCCGCTCGGCTGAGGAGGCTTAA 840
QY 841 CAGAACAGAGGCGTGGGCAAGAGGATGAAAGCTTGGGCGCAAAAGGAGCTGTGA 900
Db 841 CAGAACAGAGGCGTGGGCAAGAGGATGAAAGCTTGGGCGCAAAAGGAGCTGTGA 900
QY 901 GACCAACCCCTCAAGGGGCGAGCAACCCCACTTCAACAAGAGAGAGAAAGAAATTAACAG 960
Db 901 GACCAACCCCTCAAGGGGCGAGCAACCCCACTTCAACAAGAGAGAGAAAGAAATTAACAG 960
QY 961 ACCCATAGACCAACCCGCTTACTGTGATGAGTGTGCTTTGGCTCCGATCTGAAG 1020
Db 961 ACCCATAGACCAACCCGCTTACTGTGATGAGTGTGCTTTGGCTCCGATCTGAAG 1020
QY 1021 CGCCAGCTTGGGCGCCCGCGGATGGGAAAGGAGGATGCGCAAAAGCTCGTGTCTCTT 1080
Db 1021 CGCCAGCTTGGGCGCCCGCGGATGGGAAAGGAGGATGCGCAAAAGCTCGTGTCTCTT 1080
```

```
Db 1021 CGCCAGCTTGGGCGCCCGCGGATGGGAAAGGAGGATGCGCAAAAGCTCGTGTCTT 1080
QY 1081 GTGAGGCCACCACTTACCCCGAGGGGTAGCCACTGCGCCCGGCGCAAGGAGGACACAT 1140
Db 1081 GTGAGGCCACCACTTACCCCGAGGGGTAGCCACTGCGCCCGGCGCAAGGAGGACACAT 1140
QY 1141 GCGAGCCATTCACCCAGCTGTGCTCTCAAGACAGAGCGGAGCGCAGAGCTCCCA 1200
Db 1141 GCGAGCCATTCACCCAGCTGTGCTCTCAAGACAGAGCGGAGCGCAGAGCTCCCA 1200
QY 1201 GAAGTATATCTATGAGGTGATTACACTTTCACGCCCCCTGAAGCGGAGATTTCCATT 1260
Db 1201 GAAGTATATCTATGAGGTGATTACACTTTCACGCCCCCTGAAGCGGAGATTTCCATT 1260
QY 1261 CCTACCCACCTGAATGTCCCGAGCACTGATGATCCAGGACCAAGGAGGAGGAGGAGGAG 1320
Db 1261 CCTACCCACCTGAATGTCCCGAGCACTGATGATCCAGGACCAAGGAGGAGGAGGAGGAG 1320
QY 1321 TGGGCTCAGAGATCTCAGGCTTCAACGTCAGGAGTGAACCTTCCGAGGCGGCTGTGAC 1380
Db 1321 TGGGCTCAGAGATCTCAGGCTTCAACGTCAGGAGTGAACCTTCCGAGGCGGCTGTGAC 1380
QY 1381 TTCCAGGCTCGCTCAGTTAGCAATTCAGTCCATCTACCCAGAGAGTGGGCGCAC 1440
Db 1381 TTCCAGGCTCGCTCAGTTAGCAATTCAGTCCATCTACCCAGAGAGTGGGCGCAC 1440
QY 1441 CCAAGAACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
Db 1441 CCAAGAACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
QY 1501 GCGGAGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
Db 1501 GCGGAGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
QY 1561 TATTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
Db 1561 TATTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
QY 1621 ATCTGACATCTCTAGTGAATCTTGGCTTTTCCCGATTTGCGGATTTGGGAGGAGCAC 1680
Db 1621 ATCTGACATCTCTAGTGAATCTTGGCTTTTCCCGATTTGCGGATTTGGGAGGAGCAC 1680
QY 1681 CTCTAAGATGCTCTCTCAGAGGCTGTCAACATATCTCAAAATTAATGAGCCAGG 1740
Db 1681 CTCTAAGATGCTCTCTCAGAGGCTGTCAACATATCTCAAAATTAATGAGCCAGG 1740
QY 1741 GGCCTGACACTTCCCAATCATCTATGCTTGTGCTGCGCAAGTGGCAATTAAGCGGCTGAT 1800
Db 1741 GGCCTGACACTTCCCAATCATCTATGCTTGTGCTGCGCAAGTGGCAATTAAGCGGCTGAT 1800
QY 1801 TGGCAACTGTG 1811
Db 1801 TGGCAACTGTG 1811
```

RESULT 2
AAH16354
ID AAH16354 standard; cDNA; 1775 BP.

XX AAH16354;
AC 26-JUN-2001 (first entry)
XX Human cDNA sequence SEQ ID NO:15283.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX EP1074617-A2.
XX EP1074617-A2.
XX 07-FEB-2001.
XX

PF 28-JUN-2000; 2000EP-00116126.
 XX 29-JUN-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX (HELI-) HELIX RES INST.
 PA Ota T, Iogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T,
 XX WPI; 2001-318749/34.
 DR WPI; 2001-318749/34.
 XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.
 PS Claim 8; SEQ ID NO 15283; 2537bp + Sequence Listing; English.
 XX The present invention describes primer sets for synthesizing 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH0116 to AAH1362 and
 CC AAH1363 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
 CC represent human amino acid sequences; and AAH1362 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention.
 CC
 XX Sequence 1775 BP; 367 A; 558 C; 508 G; 342 T; 0 U; 0 Other;
 SQ
 Query Match 89.5%; Score 1621; DB 4; Length 1775;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 301 TTCACTCAGTTTGTGATCCGTAATGSAACAAATTCGAGCTTACAGTGTGT 360
 QY 551 GAGAGATTAATTAATGAACAATCTTGTAAAGCTCTTTGAGAGAGAGCTTGGAAACG 610
 Db 361 GAGAGATTAATTAATGAACAATCTTGTAAAGCTCTTTGAGAGAGAGCTTGGAAACG 420
 QY 611 GAGCTGAGCCGAGAGACACCTGTGTGACCAAGAGGACCAAGGAGCAATGAAGACCC 670
 Db 421 GAGCTGAGCCGAGAGACACCTGTGTGACCAAGAGGACCAAGGAGCAATGAAGACCC 480
 QY 671 GTGAGCTGAGCCGAG 730
 Db 481 GTGAGCTGAGCCGAG 540
 QY 731 TACCGGCTCAAGGAG 790
 Db 541 TACCGGCTCAAGGAG 600
 QY 791 ACCCGGCTCAAGGAG 850
 Db 601 ACCCGGCTCAAGGAG 660
 QY 851 GAGCTGAGCCGAG 910
 Db 661 GAGCTGAGCCGAG 720
 QY 911 TCAAGGAG 970
 Db 721 TCAAGGAG 780
 QY 971 CACACCGGCTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1030
 Db 781 CACACCGGCTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
 QY 1031 GAGGAGCCGAG 1090
 Db 841 GAGGAGCCGAG 900
 QY 1091 CCACTTACCCGAG 1150
 Db 901 CCACTTACCCGAG 960
 QY 1151 CACCGAGTGTGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1210
 Db 961 CACCGAGTGTGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 QY 1211 ATGGGTGAGTTACACTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1270
 Db 1021 ATGGGTGAGTTACACTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
 QY 1271 CTGAATGTCCGAG 1330
 Db 1081 CTGAATGTCCGAG 1140
 QY 1331 GATCTCAGGCTTCAAG 1390
 Db 1141 GATCTCAGGCTTCAAG 1200
 QY 1391 CGCTCAGTTAGCATTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1450
 Db 1201 CGCTCAGTTAGCATTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 QY 1451 AAGCCCTTGAATGATGATCTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1510
 Db 1261 AAGCCCTTGAATGATGATCTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 QY 1511 ATGGCCCTTGAATGATGATCTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1570
 Db 1321 ATGGCCCTTGAATGATGATCTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
 QY 1571 GCGCCTGTGAGGAG 1630
 Db 1381 GCGCCTGTGAGGAG 1440

QY 1631 TCTTAGTCATCTTCTGCTTTTCTGCCGATTCGAGATTGGGGGGCCACCTCTAAGATG 1690
DB 1441 TCTTAGTCATCTTCTGCTTTTCTGCCGATTCGAGATTGGGGGGCCACCTCTAAGATG 1500
QY 1691 CCTCTCCAGCCCTGTCTCAACCATCTCCAAATTAATGTCGCAACCCAGGGGCTTGCCAC 1750
DB 1501 CCTCTCCAGCCCTGTCTCAACCATCTCCAAATTAATGTCGCAACCCAGGGGCTTGCCAC 1560
QY 1751 CTCCCATATCATCATCTGTCTGTGTCGCAAGTGGGAATTAACGGCGTGAATTCGCAACCTG 1810
DB 1561 CTCCCATATCATCATCTGTCTGTGTCGCAAGTGGGAATTAACGGCGTGAATTCGCAACCTG 1620
QY 1811 G 1811
DB 1621 G 1621

RESULT 3
ABA08585
ID ABA08585 standard; cDNA; 1826 BP.
XX
AC ABA08585;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human type II procollagen homologue-encoding cDNA, SRQ ID NO:361.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haemotopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibitor; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antihastmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotropic; cardiac; virucide; antibacterial;
KW antifungal; vulnery; antilicer; ss.
XX
OS Homo sapiens.
XX
PN MO200157188-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001MO-US003800.
XX
PR 03-FEB-2000; 2000US-00496914.
XX
PR 27-APR-2000; 2000US-00560875.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI: 2001-457740/49.
XX
DR P-PDB; ABB11341.
XX
PT Human proteins and DNA encoding sequences useful for preventing, treating
XX or ameliorating a medical condition in a mammalian subject e.g. arthritis
XX and cancer.
PS Claim 1; Page 460; 1963pp; English.
CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the

CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haemotopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a cDNA encoding a
CC novel human polypeptide of the invention
XX

SQ Sequence 1826 BP; 382 A; 566 C; 515 G; 363 T; 0 U; 0 Other;

Query Match 88.9%; Score 1610.2; DB 4; Length 1826;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1612; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 197 AGTGTGGGTTCTGGGTTCTGGATTCGGGGGCGTTCAACAGTACCTGTGGGCTCC 256
DB 10 ATTCTGGGTTCTGGGTTCTGGATTCGGGGGCGTTCAACAGTACCTGTGGGCTCC 69
QY 257 TCGGGTGAATCCGTCGCGCGCGGTGCCCCGGGACGGCTTGAAGCTCCGGGG 316
DB 70 TCGGGTGAATCCGTCGCGCGCGGTGCCCCGGGACGGCTTGAAGCTCCGGGG 129
QY 317 CCCGAGCATTCGGGCTGCAATTGACGGGATCCCGGATGACCGCGGCCCGCGC 376
DB 130 CCCGAGCATTCGGGCTGCAATTGACGGGATCCCGGATGACCGCGGCCCGCGC 189
QY 377 CCTCACCGAGGGTCCAGACTGTGGGAAAGAGTGGGGGAGCGGGTCCCTGAGATCC 436
DB 190 CCTCACCGAGGGTCCAGACTGTGGGAAAGAGTGGGGGAGCGGGTCCCTGAGATCC 249
QY 437 CGATGCCCTAGCAGCCAAAGTCTCACTTATAGTGTGAACCTACATGTGACTTAC 496
DB 250 CGATGCCCTAGCAGCCAAAGTCTCACTTATAGTGTGAACCTACATGTGACTTAC 309
QY 497 TCAATTTTGTATCCGTAATAATGAGCAATTTGAAGCTACTTCAAGTGTGAGAG 556
DB 310 TCAATTTTGTATCCGTAATAATGAGCAATTTGAAGCTACTTCAAGTGTGAGAG 369
QY 557 ATTAATGAAACAATCTTTGAAGCTTTTGAAGAGGAGGAGCTGGAAAGCGGCTTG 616
DB 370 ATTAATGAAACAATCTTTGAAGCTTTTGAAGAGGAGGAGCTGGAAAGCGGCTTG 429
QY 617 GCGGAGAGACACCTGTGTACCAAGGAGACACAGGAGCATGAAGACCCCGTGAG 676
DB 430 GCGGAGAGACACCTGTGTACCAAGGAGACACAGGAGCATGAAGACCCCGTGAG 489
QY 677 CTGGCGGTAGTGGATGACAGACCTTGGGCTTCAAGACCGCTGCCAGAGTGTACCG 736
DB 490 CTGGCGGTAGTGGATGACAGACCTTGGGCTTCAAGACCGCTGCCAGAGTGTACCG 549

```

QY 737 GTCAAGCCAGGACGTATATGTGATGATGAGACTCTGTTGGAGCCAGCAGACACCCGG 796
DB 550 GTCAAGGCGAGGACGTATATGTGATGATGAGACTCTGTTGGAGCCAGCAGACACCCGG 609
QY 797 CCTACCCCAACCGGACTTGGATCCGCCCTGGGTGAGAGAGGCTTAACAGAACAGAGCGGTG 856
DB 610 CCTACCCCAACCGGACTTGGATCCGCCCTGGGTGAGAGAGGCTTAACAGAACAGAGCGGTG 669
QY 857 GGCAGAGGAGGATGAGAGGCTTGGGGGAGAGAGGAGAGGAGAGAGGAGAGGAGAGGAGAGG 916
DB 670 GGCAGAGGAGGATGAGAGGCTTGGGGGAGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGG 729
QY 917 GGCAGAGGAGGATGAGAGGCTTGGGGGAGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGG 976
DB 730 GGCAGAGGAGGATGAGAGGCTTGGGGGAGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGG 789
QY 977 CCGTCTTACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1036
DB 790 CCGTCTTACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 849
QY 1037 CCGGAGATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1096
DB 850 CCGGAGATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 909
QY 1097 ACCCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1156
DB 910 ACCCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 969
QY 1157 GCTGTCTCTTCCAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1216
DB 970 GCTGTCTCTTCCAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1029
QY 1217 GGGTTACACTCTTTCACGCCCCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1276
DB 1030 GGGTTACACTCTTTCACGCCCCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1089
QY 1277 GTCCCAAGCACTGTGATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1336
DB 1090 GTCCCAAGCACTGTGATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1149
QY 1337 AGGCCCTTCACAGTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1396
DB 1150 AGGCCCTTCACAGTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1209
QY 1397 GTTACATTTTCAATGTCATCTTACCCCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1456
DB 1210 GTTACATTTTCAATGTCATCTTACCCCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1269
QY 1457 CTTGGAATGATGATCTTTTCAATGTCATCTTACCCCAAGAGGAGGAGGAGGAGGAGGAG 1516
DB 1270 CTTGGAATGATGATCTTTTCAATGTCATCTTACCCCAAGAGGAGGAGGAGGAGGAGGAG 1329
QY 1517 CTTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1576
DB 1330 CTTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1389
QY 1577 GTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1636
DB 1390 GTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1449
QY 1637 GTTCATTTCTTGGCTTTTCTCCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1696
DB 1450 GTTCATTTCTTGGCTTTTCTCCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1509
QY 1697 TCCAGGCTTGTCTCAACCATATCTCCAAATAGTGCACACCCAGGAGGAGGAGGAGGAGG 1756
DB 1510 TCCAGGCTTGTCTCAACCATATCTCCAAATAGTGCACACCCAGGAGGAGGAGGAGGAGG 1569
QY 1757 CATCATCATTTGTCTTGTGCGCAAGTGCAGATTAACCGGCGTGATTTGCCAAGCTTGG 1811
DB 1570 CATCATCATTTGTCTTGTGCGCAAGTGCAGATTAACCGGCGTGATTTGCCAAGCTTGG 1824

```

```

RESULT 4
AAH16384
ID AAH16384 standard, cDNA, 1614 BP.
XX
AC AAH16384;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:15334.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000BP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isegai T, Nishikawa T, Hayashi K, Salto K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 15334; 2537bp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH1629 to AAH1632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 1614 BP; 317 A; 514 C; 487 G; 296 T; 0 U; 0 Other;
XX
Query Match 79.2%; Score 1434.8; DB 4; Length 1614;
Best Local Similarity 90.6%; Pred. No. 0;
Matches 1612; Conservative 0; Mismatches 2; Indels 166; Gaps 1;
QY 25 AGTACACAGCAGCAGCTGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 84
DB 1 AGTACACAGCAGCAGCTGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 60

```

QY	85	GGGCTTCAGAGACGTGAGCGCCCGCCCTGAGGAGMACTCCCCCATCCGGGGGCGCAATG	144
Db	61	GGGCTTCAGAGACGTGAGCGCCCGCCCTGAGGAGMACTCCCCCATCCGGGGGCGCAATG	120
QY	145	GTCCGGGTGCGCTGCCAGATGCTGCTGCTGCCCTCCGATTTCTGAGTGAAGATGCTGG	204
Db	121	GTCCGGGCGCGTCCGAGTGTGCTGCTGCTCCCTCGATTTCTGAGTGAAGATGCTGG	180
QY	205	GTTCTGGGTTTCTGGAATTCGCGGGCCGTTCAACATGACTGTGCTGCTGCTGCTGCTG	264
Db	181	GTTCTGGGTTTCTGGAATTCGCGGGCCGTTCAACATGACTGTGCTGCTGCTGCTGCTG	240
QY	285	GTCCCTCCGCGCGCGGTGCCCCCGGAGCGGCTTAAGCTGCGCGGAGGTCCGAGGCCCCAGG	324
Db	241	GTCCCTCCGCGCGCGGTGCCCCCGGAGCGGCTTAAGCTGCGCGGAGGTCCGAGGCCCCAGG	300
QY	325	ATTCCGGGCTGAGATTGACGGGGAATCCCGAATGCAACGCGCGCCGCCGCGCTTCAACG	384
Db	301	ATTCCGGGCTGAGATTGACGGGGAATCCCGAATGCAACGCGCGCCGCCGCGCTTCAACG	360
QY	385	ACGGGTCCAGACTTGTGTGGAAAGAAAGTGCGGGGACGGGTCCCTGAGATCCCGATGCTT	444
Db	361	ACGGGTCCAGACTTGTGTGGAAAGAAAGTGCGGGGACGGGTCCCTGAGATCCCGATGCTT	420
QY	445	ACGAGCCAGAGATGCTCAGCTTATATAGGTGTGACTTAACAATGTGACTTCACTCACTGATT	504
Db	421	ACGAGCCAGAGATGCTCAGCTTATATAGGTGTGACTTCAACAATGTGACTTCACTCACTGATT	480
QY	505	GTGATCCGTAATAATGAGACAATAATTCGAAGCTACTTCAAGTGTGTTGAGAGGATTAATG	564
Db	481	GTGATCCGTAATAATGAGACAATAATTCGAAGCTACTTCAAGTGTGTTGAGAGGATTAATG	540
QY	585	AAACATGCTTGTAAAGCTCTTTGCAAGAGGAGGCTCGGAAGCAGGGCTGTGGCCGCGAG	624
Db	541	AAACATGCTTGTAAAGCTCTTTGCAAGAGGAGGCTCGGAAGCAGGGCTGTGGCCGCGAG	600
QY	625	AGCAACCTGCTGTACCAAGGAGCAACAGGAGCAATGAAGAAGCCCGGTGAGACTGGCGGT	684
Db	601	AGCAACCTGCTGTACCAAGGAGCAACAGGAGCAATGAAGAAGCCCGGTGAGACTGGCGGT	660
QY	685	CAGTGGATGAGACCTCTCGGCTTTCAGCACCGCTGCCGAGGTGGCTACCGGATCAAGGC	744
Db	661	CAGTGGATGAGACCTCTCGGCTTTCAGCACCGCTGCCGAGGTGGCTACCGGATCAAGGC	720
QY	745	CAGACGTATATGTGATGAGACTCTGTTTGGCAGCCAGCGACGACGCCGCTTACCC	804
Db	721	CAGACGTATATGTGATGAGACTCTGTTTGGCAGCCAGCGACGACGCCGCTTACCC	780
QY	805	ACCGGACTTTCGATCCGCCCTGGGTGGAAGAGCTTAACGAACCAAGAGCTGTGGGAGAGA	864
Db	781	ACCGGACTTTCGATCCGCCCTGGGTGGAAGAGCTTAACGAACCAAGAGCTGTGGGAGAGA	840
QY	885	GGCATCGAAGGCTTGGGGGCAAAAGGAGCTGTGAGACCAACCCCTCAAGGGGAGAGAC	924
Db	841	GGCATCGAAGGCTTGGGGGCAAAAGGAGCTGTGAGACCAACCCCTCAAGGGGAGAGAC	900
QY	925	CCCCACCTTCAACCAAGAGAGAAACAATACAGACCCATCAGCTCAACCCGCTTTA	984
Db	901	CCCCACCTTCAACCAAGAGAGAAACAATACAGACCCATCAGCTCAACCCGCTTTA	960
QY	985	CTGTGATAGTGTGCTGTTTGGCTCCCGAATGGAAGGCGCAGCTTCCGGGCCCCCGGAT	1044
Db	961	CTGTGATAGTGTGCTGTTTGGCTCCCGAATGGAAGGCGCAGCTTCCGGGCCCCCGGAT	1020
QY	1045	GGCGAAGGGGATGCGCGAAGACTCGTGTCTCTTGTGAGAGCCACCACTTACCCCGAG	1104
Db	1021	GGCGAAGGGGATGCGCGAAGACTCGTGTCTCTTGTGAGAGCCACCACTTACCCCGAG	1080
QY	1105	GGGTAGCACTTGCCCGCGCCCGAGGAGACCACTGAGGCATTTCAACCACTGTGTC	1164
Db	1081	GGGTAGCACTTGCCCGCGCCCGAGGAG-----	1100
QY	1165	CTCCAAGACAGGCGGGGCGACGGCGAGACTTCCAGAAATTTATCTATGGGTGGTTACA	1224

Db	1110	-----	1109
QY	1225	CTCTTACGCGCCCTGAAAGCGGGGACATTTCCTATCCCTCACCCACCTGAATGTCCACG	1284
Db	1110	-----	1109
QY	1285	CACGTGTCATCAGCGACCAAGTGGCCCCCACAACAATAGGGCTCAGAGTCTCAGAGCTTC	1344
Db	1110	-----	1154
QY	1345	CACGTCAAGGGGTGACCTTCCCGAGCCCTCTGGTACTTCCAGGGCTCGCTCAGTTAGCAT	1404
Db	1155	CACGTCAAGGGGTGACCTTCCCGAGCCCTCTGGTACTTCCAGGGCTCGCTCAGTTAGCAT	1214
QY	1405	TTTCAGTGCATCTTACCCCAAGACGAGAGTGGGGCCACCCAGAACCAAAAGCCCCCTTGGAA	1464
Db	1215	TTTCAGTGCATCTTACCCCAAGAGTGGGGCCACCCAGAACCAAAAGCCCCCTTGGAA	1274
QY	1465	ATGATTACTCTTTCAATCAGAGGTTGGCTTAATGGGGCCACGCGCAGATATGGCCCCCTTGCCA	1524
Db	1275	ATGATTACTCTTTCAATCAGAGGTTGGCTTAATGGGGCCACGCGCAGATATGGCCCCCTTGCCA	1334
QY	1525	GGGTAGGAGGACATTCATCACCCAGGAAACCCAGTAATTAAGAAACCCCTGTGGGGGC	1584
Db	1335	GGGTAGGAGGACATTCATCACCCAGGAAACCCAGTAATTAAGAAACCCCTGTGGGGGC	1394
QY	1585	AGACAGACATATGACGGGGTGGGGCGACGTCCTCCCTTATCCGACAACTCTAGTCGATTC	1644
Db	1395	AGACAGACATATGACGGGGTGGGGCGACGTCCTCCCTTATCCGACAACTCTAGTCGATTC	1454
QY	1645	TTGGCTTTTCTCCCGCATGGATTTGGGGGGCCACCTCTAAGATGCTCTTCCAGACC	1704
Db	1455	TTGGCTTTTCTCCCGCATGGATTTGGGGGGCCACCTCTAAGATGCTCTTCCAGACC	1514
QY	1705	TGTCTCAACCATCTCAAAATTAATGTCGAACCCAGGGGCTGTGCACCTCCACATCATTC	1764
Db	1515	TGTCTCAACCATCTCAAAATTAATGTCGAACCCAGGGGCTGTGCACCTCCACATCATTC	1574
QY	1765	ATTGCTTTGCTGCGCAAGTGGGAATTAACGGCGTGATTTGCC	1804
Db	1575	ATTGCTTTGCTGCGCAAGTGGGAATTAACGGCGTGATTTGCC	1614
RESULT 5			
AD057486			
ID	AD057486	standard; cDNA; 709 BP.	
XX	AC		
XX	AD057486;		
XX	DT	29-JUN-2004 (first entry)	
XX	DE	DNA encoding hairless protein interaction partner protein #45.	
XX	XX	es; gene; human; hairless protein; Hrt; ubiquitous receptor UR; MAP1A;	
XX	KW	KIA00330 protein; monocytes antigen CD14; sphingolipid activator protein;	
XX	KW	beta-lynnulein; C11 protein; vesicle-associated membrane protein 2;	
XX	KW	aldolase A; CGI-106 protein; hypothalamus protein HSNRP1; alpha enolase;	
XX	KW	POM-2p3; quinone oxidoreductase; pumilio 1; VPS41; KIA0614 protein;	
XX	KW	splicing factor CCL4; ubiquitin; beta-mannosidase; hair growth.	
XX	OS	Homo sapiens.	
XX	PN	US2004086945-A1.	
XX	PD	06-MAY-2004.	
XX	PF	02-JUN-2003; 2003US-00452858.	
XX	PR	03-JUN-2002; 2002US-0385414P.	
XX	RA	(PROC) PROCTER & GAMBLE CO.	
XX	XX		

P1 Sreekrishna K, Gerwe GS, Toerner DR;
XX WPI: 2004-430095/40.
DR P-PSDB; AD057487.
XX
PT New composition comprising mouse truncated hairless protein-human
PT interacting partner protein or nucleic acid complexes, useful for
PT screening test compounds that inhibit or enhance hair growth.
XX
PS Claim 6; SEQ ID NO 78; 60pp; English.
XX
CC The invention relates to a composition comprising a mouse truncated
CC hairless (Hr) protein-human interacting partner protein or nucleic acid
CC complex. The human interacting partner protein comprises a molecule
CC selected from ubiquitous receptor UR, MAP1A, KIAA0930 protein, monocytes
CC antigen CD14, sphingosin lipid activator protein, beta-synuclein, C11
CC protein, vesicle-associated membrane protein 2, aldolase A, CGI-106
CC protein, hypohalimus protein HSNMP1, alpha enolase, POM-2P3, quinone
CC oxidoreductase, pumilio 1, VPS41, KIAA0614 protein, splicing factor
CC CCL4, ubiquitin, beta-mannosidase. Also described are: methods of
CC assaying a test compound for agonist or antagonist activity for the above
CC composition; and methods of inhibiting or increasing hair growth on a
CC surface in a subject. The present sequence represents DNA encoding a
CC hairless protein human interacting partner used in the composition.
XX
SQ Sequence 709 BP; 146 A; 247 C; 203 G; 110 T; 0 U; 3 Other;

Query Match 38.0%; Score 688.2; DB 12; Length 709;
Best Local Similarity 99.1%; Pred. No. 8.7e-176;
Matches 690; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 682 CGTCAGTGGATGAGACGCTCGGCTTCAGACCGCTGCGAGGTGCTACCGGGCTCA 741
DB 14 CGTCAGCGGATGACGACCTTCGCTTCAGACCGCTGCGAGGTGCTACCGGGCTCA 73
QY 742 GGGCAGGAGCTCATATGTGATGAGACTCTGTTTGGCAGCCGAGGACACCGGCTTAC 801
DB 74 GGGCAGGAGCTCATATGTGATGAGACTCTGTTTGGCAGCCGAGGACACCGGCTTAC 133
QY 802 CCCACCGGATCTTCATCGGCTCGGCTGGTGGAGAGGCTTAAACAGACGAGGCTGGGCA 861
DB 134 CCCACCGGATCTTCATCGGCTCGGCTGGTGGAGAGGCTTAAACAGACGAGGCTGGGCA 193
QY 862 GGAAGCATGGAAGGCTTGGGGGCAAGGGGAGCTGTAGACCCCTCAAGGGGGAG 921
DB 194 GGAAGCATGGAAGGCTTGGGGGCAAGGGGAGCTGTAGACCCCTCAAGGGGGAG 253
QY 922 CACCCGCCCTTCACACCAAGAGAGAAATATACAGCCATCAGCCACACCCGCTC 981
DB 254 CACCCGCCCTTCACACCAAGAGAGAAATATACAGCCATCAGCCACACCCGCTC 313
QY 982 TTACTGTGATGATGCTGCTGTTTGGCTCCGATCTGAGAGGCGCAGCTTGGGGGCGCG 1041
DB 314 TTACTGTGATGATGCTGCTGTTTGGCTCCGATCTGAGAGGCGCAGCTTGGGGGCGCG 373
QY 1042 GATGCGGAAGGGGGAGTGGCGAAAGCTCGTCTCTTGTGAGCGCAACCACTTACCC 1101
DB 374 GATGCGGAAGGGGGAGTGGCGAAAGCTCGTCTCTTGTGAGCGCAACCACTTACCC 433
QY 1102 CAGGGGTAGCCACTCGCCCGCCCGCAGGAGGAGCACTCGGAGGACTTTCACCCAGCTGG 1161
DB 434 CAGGGGTAGCCACTCGCCCGCCCGCAGGAGGAGCACTCGGAGGACTTTCACCCAGCTGG 493
QY 1162 TCCTTCAGAGCAGAGCGGGGGCCAGCGGAGACTCCGAGAGTTATTTATGGTGGTT 1221
DB 494 TCCTTCAGAGCAGAGCGGGGGCCAGCGGAGACTCCGAGAGTTATTTATGGTGGTT 553
QY 1222 ACATCTTTCAGCGCCCTCGAAGCGGGGACTTTCATTCCTCCACCCAGCTGAATGTCC 1281
DB 554 ACATCTTTCAGCGCCCTCGAAGCGGGGACTTTCATTCCTCCACCCAGCTGAATGTCC 613
QY 1282 CAGCACTGTGATCAGGACCAAGTGGCCCGCAGCAAAATGGGCTCAGATCTCAGGCC 1341

DB 614 CAGCACTGTGATCAGGACCAAGTGGCCCGCCAGCAAAATGGGCTCAGATCTCAGGCC 673
QY 1342 TTTCACGTCAGGGGAGTACCTTCCGAGAGCCCTCGGT 1377
DB 674 TTTCACGTCAGGGGAGTACCTTCCGAGAGCCCTCGGT 709

RESULT 6
AAH07100
ID AAH07100 standard; cDNA; 784 BP.
XX
AC AAH07100;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:3935.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
FN EPI074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUN-2000; 2000EP-00116126.
XX
PR 29-JUN-1999; 99JP-00248036.
XX
PR 27-AUG-1999; 99JP-00300253.
XX
PR 11-JAN-2000; 2000JP-00118776.
XX
PR 02-MAY-2000; 2000JP-00183767.
XX
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 1; SEQ ID NO 3935; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dr primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification; where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 784 BP; 174 A; 224 C; 251 G; 130 T; 0 U; 5 Other;

Query Match 36.5%; Score 661.4; DB 4; Length 784;

CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention

Query Match 31.0%; Score 562; DB 4; Length 597;
 Best Local Similarity 97.1%; Pred. No. 1.2e-141;
 Matches 579; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

25 AGTACACACGACCTGATGAGTGGACACACAGAGACCTCTTCAGTTTAAAGACCTCT 84
 1 AGTACACACGACCTGATGAGTGGACACACAGAGACCTCTTCAGTTTAAAGACCTCT 60

85 GGGCTCTAGAGACGTCGCGCCGCCCTTGGGCGGACCTCCCATCCGCGGCGGAAATG 144
 61 GGGCTCTAGAGACGTCGCGCCGCCCTTGGGCGGACCTCCCATCCGCGGCGGAAATG 120

145 GTCCGGGTCGCGTCGCGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 204
 121 GTCCGGGTCGCGTCGCGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 180

205 GTTCTGGGTTTCTGATTCGCGGCGGCGGCTTACACAGTACCTGTGCTGCTGCTGCTG 264
 181 GTTCTGGGTTTCTGATTCGCGGCGGCGGCTTACACAGTACCTGTGCTGCTGCTGCTG 240

265 GTCCGTCGCGCGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 324
 241 GTCCGTCGCGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 300

325 ATTCCGGGCTGACAGTTGACGCGGATCCCGGATGACCGCGGCGGCGGCGGCGGCGGCG 384
 301 ATTCCGGGCTGACAGTTGACGCGGATCCCGGATGACCGCGGCGGCGGCGGCGGCGGCG 360

385 ACGGTCACAGACCTGTCGGAAGAGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 444
 361 ACGGTCACAGACCTGTCGGAAGAGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 420

445 ACGAGCCAAAGATGCTCAGCTTTATATGATGATGATGATGATGATGATGATGATGATGAT 504
 421 ACGAGCCAAAGATGCTCAGCTTTATATGATGATGATGATGATGATGATGATGATGATGAT 480

505 GTGATCCGTAATAATGACCAATTCGAACTACTTAC-AGTCTGTTGAGAGATTAAT 563
 481 GTGATCCGTAATAATGACCAATTCGAACTACTTAC-AGTCTGTTGAGAGATTAAT 540

564 GAAACATGCTTTGTAAGCTTTTGCAGAGAGAGCTCGAAGAGGCGCTGCGC 619
 541 GAAACATGCTTTGTAAGCTTTTGCAGAGAGAGCTCGAAGAGGCGCTGCGC 596

RESULT 9
 ADF82731
 ID ADF82731 standard; DNA; 816 BP.

AC ADF82731;
 DT 26-FEB-2004 (first entry)

XX Leukaemia-related DNA sequence #3287.
 XX Cytostatic; Gene therapy; leukaemia; ss.
 XX Unidentified.
 XX W02003039443-A2.
 XX 15-MAY-2003.
 XX 04-NOV-2002; 2002WO-EP012303.

CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention

Query Match 31.0%; Score 562; DB 4; Length 597;
 Best Local Similarity 97.1%; Pred. No. 1.2e-141;
 Matches 579; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

25 AGTACACACGACCTGATGAGTGGACACACAGAGACCTCTTCAGTTTAAAGACCTCT 84
 1 AGTACACACGACCTGATGAGTGGACACACAGAGACCTCTTCAGTTTAAAGACCTCT 60

85 GGGCTCTAGAGACGTCGCGCCGCCCTTGGGCGGACCTCCCATCCGCGGCGGAAATG 144
 61 GGGCTCTAGAGACGTCGCGCCGCCCTTGGGCGGACCTCCCATCCGCGGCGGAAATG 120

145 GTCCGGGTCGCGTCGCGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 204
 121 GTCCGGGTCGCGTCGCGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 180

205 GTTCTGGGTTTCTGATTCGCGGCGGCGGCTTACACAGTACCTGTGCTGCTGCTGCTG 264
 181 GTTCTGGGTTTCTGATTCGCGGCGGCGGCTTACACAGTACCTGTGCTGCTGCTGCTG 240

265 GTCCGTCGCGCGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 324
 241 GTCCGTCGCGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 300

325 ATTCCGGGCTGACAGTTGACGCGGATCCCGGATGACCGCGGCGGCGGCGGCGGCGGCG 384
 301 ATTCCGGGCTGACAGTTGACGCGGATCCCGGATGACCGCGGCGGCGGCGGCGGCGGCG 360

385 ACGGTCACAGACCTGTCGGAAGAGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 444
 361 ACGGTCACAGACCTGTCGGAAGAGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 420

445 ACGAGCCAAAGATGCTCAGCTTTATATGATGATGATGATGATGATGATGATGATGATGAT 504
 421 ACGAGCCAAAGATGCTCAGCTTTATATGATGATGATGATGATGATGATGATGATGATGAT 480

505 GTGATCCGTAATAATGACCAATTCGAACTACTTAC-AGTCTGTTGAGAGATTAAT 563
 481 GTGATCCGTAATAATGACCAATTCGAACTACTTAC-AGTCTGTTGAGAGATTAAT 540

564 GAAACATGCTTTGTAAGCTTTTGCAGAGAGAGCTCGAAGAGGCGCTGCGC 619
 541 GAAACATGCTTTGTAAGCTTTTGCAGAGAGAGCTCGAAGAGGCGCTGCGC 596

RESULT 9
 ADF82731
 ID ADF82731 standard; DNA; 816 BP.

AC ADF82731;
 DT 26-FEB-2004 (first entry)

XX Leukaemia-related DNA sequence #3287.
 XX Cytostatic; Gene therapy; leukaemia; ss.
 XX Unidentified.
 XX W02003039443-A2.
 XX 15-MAY-2003.
 XX 04-NOV-2002; 2002WO-EP012303.


```
Db 497 CCCCCCANAAATGGGCTGAGATNTCAGGCTTCCACGTCAGGGGGAAGCTTCCGGA 438
Qy 1368 GCGCCCTGTGACTTCCAGGGGCTGGCTAGTTCAGTTCATTCATCCCAAGAC 1427
Db 437 GCGCCCTGTGACTTCCAGGGGCTGGCTAGTTCAGTTCATTCATCCCAAGAC 378
Qy 1428 GAGGTGGGGCCACCCAGAAACCAAGCCCTTGGAAATGATCTTTCATCAGGGTTG 1487
Db 377 GAGGTGGGGCCACCCAGAAACCAAGCCCTTGGAAATGATCTTTCATCAGGGTTG 318
Qy 1488 CCTATGGGGCCACCGGCGACAGGTATGGCCCTTGGCAGGGTGAAGACATTCATCACCC 1547
Db 317 CCTATGGGGCCACCGGCGACAGGTATGGCCCTTGGCAGGGTGAAGACATTCATCACCC 258
Qy 1548 AGGGAACCCCGAGTATTAAAGAGCCCTGTGGGGCGACAGACATAGACAGGGGTGGGC 1607
Db 257 AGGGAACCCCGAGTATTAAAGAGCCCTGTGGGGCGACAGACATAGACAGGGGTGGGC 198
Qy 1608 AGTGGCTCCCTTATCTGACATATCTAGTGGATTCTTGGCTTTTCCCGGATTTGGGG 1667
Db 197 AGTGGCTCCCTTATCTGACATATCTAGTGGATTCTTGGCTTTTCCCGGATTTGGGG 138
Qy 1668 ATTTGGGGGCGACCTCTAAGATGCTCTCTCCAGCCCTGTCTCAACCATCTCAAAATTA 1727
Db 137 ATTTGGGGGCGACCTCTAAGATGCTCTCTCCAGCCCTGTCTCAACCATCTCAAAATTA 78
Qy 1728 GTGCCAACCCGAGGGGCTGTGGCACTCTCCACATCATCATCTCTTGTGGCCAGTGGCA 1787
Db 77 GTGCCAACCCGAGGGGCTGTGGCACTCTCCACATCATCATCTCTTGTGGCCAGTGGCA 18
Qy 1788 TAAACGGCGGTGATTGCC 1804
Db 17 TAAACGGCGGTGATTGCC 1
```

RESULT 11

AD116303/c
ID AD116303 standard; DNA; 3673 BP.

AC AD116303;

DT 22-APR-2004 (first entry)

DE Human nucleic acid-associated protein (NAAP) coding sequence #38.

XX human; nucleic acid-associated protein; NAAP; autoimmune disorder;
XX inflammatory disorder; AIDS; allergy; infection; metabolic disorder;
XX obesity; reproductive disorder; infertility; neurological disorder;
XX Parkinson's disease; Alzheimer's disease; cardiovascular disorder;
XX myocardial infarction; hypertension; eye disorder;
XX cell proliferative disease; cancer; ds; gene.

OS Homo sapiens.

XX MO2003094848-A2.

PN 20-NOV-2003.

PF 09-MAY-2003; 2003WO-US014450.

XX 10-MAY-2002; 2002US-0379843P.

PR 24-MAY-2002; 2002US-0383457P.

PR 31-MAY-2002; 2002US-0384699P.

PR 06-JUN-2002; 2002US-0387265P.

XX (INCY-) INCYTE CORP.

XX Kable AE, Elliott VS, Tran UK, Ramkumar J, Margulis JP, Chawla NK;
PI Richardson TM, Bulloch SA, Khare R, Lee SY, Lai PG, Tang YT, Yue H;
PI Swarnakar A, Becha SD, Hafalia AJA, Chang H, Baughn MR, Borowsky ML;
PI Gietzen KJ, He A, Forsythe J, Sprague WM, Blake JT, Warren BA;
PI Mason PM, Ison CH, Lindquist BA, Wilson AD, Jin P;

XX MPI: 2004-011999/01.
DR P-PSDB: AD116253.
XX
PT New human nucleic acid associated proteins and polynucleotides, useful
PT for diagnosing, preventing or treating diseases or conditions associated
PT with aberrant protein expression, e.g. cancer, AIDS, atherosclerosis or
PT stroke.
XX
PS Claim 5; SEQ ID NO 88; 400bp; English.

XX The invention comprises the amino acid and coding sequences of human
CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
CC the invention are useful in diagnosing, preventing and treating
CC diseases/conditions associated with altered expression of NAAP, such as:
CC autoimmune/inflammatory disorders (e.g. AIDS and allergies), infections
CC (e.g. bacterial and viral), metabolic disorders (e.g. obesity),
CC reproductive disorders (e.g. infertility), neurological disorders (e.g.
CC Parkinson's disease and Alzheimer's disease), cardiovascular disorders
CC (e.g. myocardial infarction and hypertension), eye disorders, or cell
CC proliferative diseases (e.g. cancer). The present DNA sequence encodes a
CC human NAAP protein of the invention.

XX Sequence 3673 BP; 785 A; 1129 C; 1196 G; 563 T; 0 U; 0 Other;

Query Match 26.6%; Score 481.4; DB 12; Length 3673;

Best Local Similarity 99.6%; Pred. No. 1.7e-119; Matches 493; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
Qy 1 AGTGGCTCCGCGCCCTGCGGGGCTAGTACACAGCACCTGAGTGAAGGACAGAGAC 60
Db 495 AGTGGCTCCGCGCCCTGCGGGGCTAGTACACAGCACCTGAGTGAAGGACAGAGAC 436
Qy 61 CCTTCATGTTTAAAGGACCTCTGAGGCTTGAAGAGCGTGGCGCCGCTTGGCGGAC 120
Db 435 CCTTCATGTTTAAAGGACCTCTGAGGCTTGAAGAGCGTGGCGCCGCTTGGCGGAC 376
Qy 121 TCCCGCCATCCGCGGGCGGAAATGTCGGGTCGGCTCCGAGTCTCTGCTGCTCCC 180
Db 375 TCCCGCCATCCGCGGGCGGAAATGTCGGGTCGGCTCCGAGTCTCTGCTGCTCCC 316
Qy 181 TGGTTGCTGGGTGCAAGATGCTGGGTTCTGGGTTTCTGGATTTCGGGGCGGTTACAGCT 240
Db 315 TGGTTGCTGGGTGCAAGATGCTGGGTTCTGGGTTTCTGGATTTCGGGGCGGTTACAGCT 256
Qy 241 AGCTGTGCGCGGCTCTCTGAGTGAATCCGTCGCGCGCGGTGCGCCGCGGACGAGCTTAGGC 300
Db 255 AGCTGTGCGCGGCTCTCTGAGTGAATCCGTCGCGCGCGGTGCGCCGCGGACGAGCTTAGGC 196
Qy 301 TGCCTGGGGGTCCGGGGCCCGACGATTCGGGGCTGACATTGACGGGGATCCGGATGCA 360
Db 195 TGCCTGGGGGTCCGGGGCCCGACGATTCGGGGCTGACATTGACGGGGATCCGGATGCA 136
Qy 361 CCGCGCGCGCCCGCGGCGCTTACCGAGC-GGGTCACAGACCTGTGGGAAGAAGGTGGGGGA 419
Db 135 CCGCGCGCGCCCGCGGCGCTTACCGAGC-GGGTCACAGACCTGTGGGAAGAAGGTGGGGGA 76
Qy 420 CGGGTCCCTGAGATCCCGATGCTTACGAGGCAAGATGCTTATAGGTGACCT 479
Db 75 CGGGTCCCTGAGATCCCGATGCTTACGAGGCAAGATGCTTATAGGTGACCT 16
Qy 480 ACACATGTGACTTCA 494
Db 15 ACACATGTGACTTCA 1
```

RESULT 12

ID ACH14455 standard; cDNA; 477 BP.

AC ACH14455;

DT 13-OCT-2003 (first entry)

XX Human adult brain cDNA #1667.
 DE Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
 XX genome mapping; biodiversity; genetic disorder.
 KW Homo sapiens.
 OS US2003073623-A1.
 XX
 PN 17-APR-2003.
 PD 30-JUL-2001; 2001US-00918995.
 XX
 PF 30-JUL-2001; 2001US-00918995.
 XX
 PR 30-JUL-2001; 2001US-00918995.
 XX
 PA (DRMA/) DRMANAC R T.
 PA (LABA/) LABAT I.
 PA (STAC/) STACHE-CRAIN B.
 PA (DICK/) DICKSON M C.
 PA (JONE/) JONES L W.
 XX
 PI Dmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
 DR WPI; 2003-615964/58.
 XX
 DR New polynucleotide sequences obtained from various cDNA libraries, useful
 PT as hybridization probes, as oligomers for PCR, for chromosome and gene
 PT mapping, in the recombinant production of protein, or in generating
 PT antisense DNA or RNA.
 XX
 PS Claim 1; SEQ ID NO 1667; 44p; English.
 XX
 CC The invention relates to an isolated polynucleotide comprising any one of
 CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
 CC determined by the technique of SBH (sequencing by hybridisation). Also
 CC included is a purified polypeptide comprising a sequence corresponding to
 CC a reading frame of the novel polynucleotide. The nucleic acid sequences
 CC are useful in diagnostics as expressed sequence tags (EST) for
 CC identifying expressed genes or for physical mapping of the human genome,
 CC in forensic, in assessing biodiversity, or in identifying mutations
 CC responsible for genetic disorders and other traits. The nucleotide
 CC sequences are also useful as hybridisation probes, as oligomers for PCR,
 CC for chromosome and gene mapping, in the recombinant production of
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide
 CC is useful for generating antibodies specific for it. The present sequence
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
 CC obtained in electronic format directly from the printed specification, but was
 CC obtained in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?docid=20030073623
 XX
 SQ Sequence 477 BP; 86 A; 174 C; 125 G; 90 T; 0 U; 2 Other;
 XX
 Query Match 24.1%; Score 436.6; DB 9; Length 477;
 Best Local Similarity 99.1%; Pred. No. 1e-107;
 Matches 439; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 XX
 QY 986 TGGAGAGGCTGCTGTTGGCTCCCATCGAAGGCGCCAGCTTCGGGGCCCGCGGATG 1045
 DB 35 TGGAAATTCGTCGCTGTTGGCTCCCATCGAAGGCGCCAGCTTCGGGGCCCGCGGATG 94
 QY 1046 GCGAAGGGGATCCCGAAGACTCCGTCCTCTTGTGAGCGCACCACTTACCCAGG 1105
 DB 95 GCGAAGGGGATCCCGAAGACTCCGTCCTCTTGTGAGCGCACCACTTACCCAGG 154
 QY 1106 GGTAGCCATCGCCCGCCCGAGGAGGACCACTGCGAGCACTTACCCAGGTTGCC 1165
 DB 155 GTAGGCACTCGCCCGCCCGAGGAGGACCACTGCGAGCACTTACCCAGGTTGCC 214
 QY 1166 TCCAAAGACAGAGCGGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1225
 DB 215 TCCAAAGACAGAGCGGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 274

QY 1226 TCTTCAAGCCCGCTGAGAGGGGAGACTTCCATTCCCTGACCCAGCTGAATGCCAGC 1285
 DB 275 TCTTCAAGCCCGCTGAGAGGGGAGACTTCCATTCCCTGACCCAGCTGAATGCCAGC 334
 QY 1286 ACTGGTCATTCAGACCAACAGTGGCCCCCAACAATAGGCTTACAGATCTCAGGCTTCC 1345
 DB 335 ACTGGTCATTCAGACCAACAGTGGCCCCCAACAATAGGCTTACAGATCTCAGGCTTCC 394
 QY 1346 ACCTCAGGGGTGACCTTCCGAGAGCCCCCTGGTGAACCTTCAGAGGCTGCTCAGTTACATT 1405
 DB 395 ACCTCAGGGGTGACCTTCCGAGAGCCCCCTGGTGAACCTTCAGAGGCTGCTCAGTTACATT 454
 QY 1406 TCAGTCCATCTACCCACGACG 1428
 DB 455 TCAGTCCATCTACCCACGACG 477
 XX
 RESULT 13
 ACH3954
 ID ACH43954 standard; cDNA; 484 BP.
 XX
 AC ACH43954;
 XX
 DT 13-OCT-2003 (first entry)
 XX
 DE Human foetal brain cDNA #4679.
 XX
 KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
 KW genome mapping; biodiversity; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN US2003073623-A1.
 PD 17-APR-2003.
 XX
 PF 30-JUL-2001; 2001US-00918995.
 XX
 PR 30-JUL-2001; 2001US-00918995.
 XX
 PA (DRMA/) DRMANAC R T.
 PA (LABA/) LABAT I.
 PA (STAC/) STACHE-CRAIN B.
 PA (DICK/) DICKSON M C.
 PA (JONE/) JONES L W.
 XX
 PI Dmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
 XX
 DR WPI; 2003-615964/58.
 XX
 DR New polynucleotide sequences obtained from various cDNA libraries, useful
 PT as hybridization probes, as oligomers for PCR, for chromosome and gene
 PT mapping, in the recombinant production of protein, or in generating
 PT antisense DNA or RNA.
 XX
 PS Claim 1; SEQ ID NO 31166; 44p; English.
 XX
 CC The invention relates to an isolated polynucleotide comprising any one of
 CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
 CC determined by the technique of SBH (sequencing by hybridisation). Also
 CC included is a purified polypeptide comprising a sequence corresponding to
 CC a reading frame of the novel polynucleotide. The nucleic acid sequences
 CC are useful in diagnostics as expressed sequence tags (EST) for
 CC identifying expressed genes or for physical mapping of the human genome,
 CC in forensic, in assessing biodiversity, or in identifying mutations
 CC responsible for genetic disorders and other traits. The nucleotide
 CC sequences are also useful as hybridisation probes, as oligomers for PCR,
 CC for chromosome and gene mapping, in the recombinant production of
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide
 CC is useful for generating antibodies specific for it. The present sequence
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
 CC obtained in electronic format directly from the printed specification, but was
 CC obtained in electronic format directly from USPTO at

CC seqdata.uniprot.org/sequence.html?docID=20030073623
 XX Sequence 484 BP; 118 A; 129 C; 137 G; 85 T; 0 U; 15 Other;
 SQ Query Match 22.7%; Score 410.6; DB 9; Length 484;
 Best Local Similarity 93.9%; Pred. No. 1.1e-100;
 Matches 419; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 426 CCGTAGAGTCCCGATGCTGAGCCAGAGGCTGAGCTTTATAGGTGACCTACAT 485
 DB 39 CCGTAGAGTCCCGATGCTGAGCCAGAGGCTGAGCTTTATAGGTGACCTACAT 98
 QY 486 GTGACTTACCTGAGCTTTGATGATCCGTAATGAGCAAAATTCGAGGCTACTTCACAGTG 545
 DB 99 GTGACTTACCTGAGCTTTGATGATCCGTAATGAGCAAAAGGCGAAGCTACTTCACAGAG 158
 QY 546 CTGTTGAGAGATTAATGAAAACAATGCTTTGTAAGCTTTTTCAGAGAGGAGCTTCGGA 605
 DB 159 CTGTTGAGAGATTAATGAAAACAATGCTTTGTAAGCTTTTTCAGAGAGGAGCTTCGGA 218
 QY 606 AGCAGAGGCTGCGCCGCGAGAGACACCTGCTGACACAGGACCAAGGACGATGAGA 665
 DB 219 AGCAGAGGCTGCGCCGCGAGAGACACCTGCTGACACAGGACCAAGGACGATGAGA 278
 QY 666 CCCCCGTGAGCTGCGCCGCTGAGTGGAGTGCAGACCTCGGCTTCAGACGCTGCGAG 725
 DB 279 CCCCCGTGAGCTGCGCCGCTGAGTGGAGTGCAGACCTCGGCTTCAGACGCTGCGAG 338
 QY 726 GTGCTTACCGGGTCAAGGCGCAGAGCGTCAATGTAATGAGACTGTGTTTGGCAGCCGAG 785
 DB 339 GTGCTTACCGGGTCAAGGCGCAGAGCGTCAATGTAATGAGACTGTGTTTGGCAGCCGAG 398
 QY 786 CAGGACACCCGCTCTACCCCACTTCGATCCGCTGCGGTGAGAGAGGCTAACAGAA 845
 DB 399 CAGGACACCCGCTCTACCCCACTTCGATCCGCTGCGGTGAGAGAGGCTAACAGAA 458
 QY 846 CCAAGAGCGTGGGCAAGAGGAGCATCG 871
 DB 459 CCAAGAGCGTGGGCAAGAGGAGCATCG 484

RESULT 14
 ADF81362
 ID ADF81362 standard; DNA; 572 BP.
 AC ADF81362;
 AC 26-FEB-2004 (first entry)
 DT 26-FEB-2004 (first entry)
 XX Leukaemia-related DNA sequence #1918.
 DE Leukaemia-related DNA sequence #1918.
 XX Cytostatic; Gene therapy; leukaemia; ss.
 KM Cytostatic; Gene therapy; leukaemia; ss.
 XX Unidentified.
 OS Unidentified.
 XX WO2003039443-A2.
 PN WO2003039443-A2.
 XX 15-MAY-2003.
 PD 15-MAY-2003.
 PF 04-NOV-2002; 2002MO-EP012303.
 XX 04-NOV-2002; 2002MO-EP012303.
 PR 05-NOV-2001; 2001EP-00126244.
 XX 05-NOV-2001; 2001EP-00126244.
 PR 30-APR-2002; 2002EP-00009758.
 XX 30-APR-2002; 2002EP-00009758.
 XX (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
 PA (UYLU-) UNIV LUDWIG MAXIMILIANS.
 PA (HAPE/) HAFERLACH T.
 PA (SCHO/) SCHOCH C.
 PA (KERN/) KERN W.
 XX HAFERLACH T, SCHOCH C, KERN W, Kohlmann A, Schittger S, Dugas M;
 PI Elle R, Broers B, Mergenthaler S;
 XX

DR WPI; 2003-505037/47.
 XX Determining the subtype of leukemia cells and whether a patient sample
 PT contains leukemia cells or other cells, useful for treating leukemia,
 PT comprises determining the expression profile of a group of markers in a
 PT patient sample.
 XX Disclosure; SEQ ID NO 1918; 2938bp; English.
 PS The present invention relates to a method (M1) for determining the
 CC subtype of leukemia cells and whether a patient sample contains
 CC leukemia cells. The method comprises determining the expression profile
 CC of a group of markers in a patient sample. The method is useful for
 CC determining the presence of leukemia cells, its types or subtypes, and
 CC for the preparation of a medicament for treating leukaemia.

SQ Sequence 572 BP; 113 A; 154 C; 129 G; 126 T; 0 U; 50 Other;
 Query Match 20.4%; Score 369; DB 10; Length 572;
 Best Local Similarity 94.9%; Pred. No. 2.3e-89;
 Matches 369; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1397 GTTAGCATTTGAGTCCATCTACCCACAGAGGTGGGCGCACCCAGAAAACAAAGCCC 1456
 DB 1 GTTAGCATTTGAGTCCATCTACCCACAGAGGTGGGCGCACCCAGAAAACAAAGCCC 60
 QY 1457 CTTGGAAATATATCTCTTTTATCAGGGTTCCTATGGGGCGACGCGCATAGTATGACC 1516
 DB 61 CTTGGAAATATATCTCTTTTATCAGGGTTCCTATGGGGCGACGCGCATAGTATGACC 120
 QY 1517 CTTGGCAAGGTGAGAGAGACATTCATCACCCAGGAAACCCAGGTATTAAGAAACCCCT 1576
 DB 121 CTTGGCAAGGTGAGAGAGACATTCATCACCCAGGAAACCCAGGTATTAAGAAACCCCT 180
 QY 1577 GTGGGGGAGACAGACATAGCAGAGGTGGGAGTGCCTCTTATCTGACAACTCTTA 1636
 DB 181 GTGGGGGAGACAGACATAGCAGAGGTGGGAGTGCCTCTTATCTGACAACTCTTA 240
 QY 1637 GTGCATTTTGGCTTTTCTCCGATTCGGGATTTGGGGGCGACCTCTAAGATGCTCTC 1696
 DB 241 GTGCATTTTGGCTTTTCTCCGATTCGGGATTTGGGGGCGACCTCTAAGATGCTCTC 300
 QY 1697 TCCAGCCCTGTCTAACAATCTCAATTAATGAGCAACCAAGGGGCTGGACCTCCCA 1756
 DB 301 TCCAGCCCTGTCTAACAATCTCAATTAATGAGCAACCAAGGGGCTGGACCTCCCA 360
 QY 1757 CATCATCATTTGCTTGTGCTGCAAGTGGC 1785
 DB 361 CATCATCATTTGCTTGTGCTGCAAGTGGC 389

RESULT 15
 AAH12103/C
 ID AAH12103 standard; cDNA; 578 BP.
 AC AAH12103;
 AC 26-JUN-2001 (first entry)
 DT 26-JUN-2001 (first entry)
 XX Human cDNA clone (3'-primer) SEQ ID NO:8938.
 DE Human cDNA clone (3'-primer) SEQ ID NO:8938.
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 KM Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX Homo sapiens.
 OS Homo sapiens.
 XX EP1074617-A2.
 PN EP1074617-A2.
 XX 07-FEB-2001.
 PD 07-FEB-2001.
 PF 28-JUL-2000; 2000EP-00116126.
 XX 28-JUL-2000; 2000EP-00116126.
 PR 29-JUL-1999; 99JP-00248036.
 XX 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 XX 27-AUG-1999; 99JP-00300253.
 XX

PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.

PS Claim 3; SEQ ID NO 8938; 2537bp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dr primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention

XX Sequence 578 BP; 134 A; 142 C; 168 G; 127 T; 0 U; 7 Other;

Query Match 18.9%; Score 341.8; DB 4; Length 578;
Best Local Similarity 93.6%; Pred. No. 5.3e-82;
Matches 396; Conservative 0; Mismatches 23; Indels 4; Gaps 4;

QY 1392 GCTCAGTACGATTTTCAGTGCATCTACCCACAGACGAGTGGGCCACCCAGAACCAA 1451
DB 576 GCTCAGTACGATTTTCAGTGCATCTACCCACAGACGAGTGGGCCACCCAGAACCAA 517
QY 1452 AGCCCCCTTGAATGATATCATCTTTTCATCA-GGGTTGCTATGGGGCCAGGGCAGGT 1510
DB 516 AGCCCCCTTGAATGATATCATCTTTTCATCA-GGGTTGCTATGGGGCCAGGGCAGGT 457
QY 1511 ATGGCCCTTGGC-AGGGTAGAGA-GA-CATTATCACCAGGAAACCCAGTATTAAAG 1568
DB 456 TAGGCCCTTGGCAGAGGTAGAGACCATTTATCACCAGGAAACCCAGTATTAAAG 397
QY 1569 AAGCCCTTGGGGGAGAGACATATAGCAGGGGTGGGGTGGCTTCTTATCTGAC 1628
DB 396 AAGCCCTTGGGGGAGAGACATATAGCAGGGGTGGGGTGGCTTCTTATCTGAC 338
QY 1629 AATCTAGTCGATTTCTGCTTTTCTCCGATTTGGCGATTTGGGGGACCACTTAA 1688
DB 337 AATCTAGTCGATTTCTGCTTTTCTCCGATTTGGCGATTTGGGGGACCACTTAA 278
QY 1689 TGCCTCTTCCAGCCCTGTCTCAACCATCTCCAAATTAAGTGCACCCAGGGGCTTGGC 1748
DB 277 TGCCTCTTCCAGCCCTGTCTCAACCATCTCCAAATTAAGTGCACCCAGGGGCTTGGC 218
QY 1749 ACCTCCCATCATTCATTTGCTTGTGCGCAAGTGGGAATAAACGGGTGATTGCCAAC 1808

DB 217 ACCTCCCATCATTCATTTGCTTGTGCGCAAGTGGGAATAAACGGGTGATTGCCAAC 158
QY 1809 TGG 1811
DB 157 TGG 155

Search completed: March 28, 2005, 01:58:00
Job time : 1020 secs

;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FASTSEQ for Windows Version 4.0
;; SEQ ID NO 16749
;; LENGTH: 18351
;; TYPE: DNA
;; ORGANISM: Human
US-09-949-016-16749

Query Match
Best Local Similarity 70.2%; Score 54.4; DB 4; Length 18351;
Pred. No. 0.0003;
Matches 73; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 487 TGACCTCAGCTGTTTGATCCGTTAAATGAGCAAAATTCGAACTTCACTGTC 546
DB 13712 TCACTGAGCTTCTTTCTCATTTGTTAAATGAGGCTATTAGACCTCACTCAAGAGC 13771
QY 547 TGTGAGAGATTAAATGAACAATGCTTTGTAAGCTCTTGGCA 590
DB 13772 TGTGTAAGATGAATGAATCAATACATGTAAGTTCTTAACA 13815

RESULT 3
US-09-949-016-16287/c
;; Sequence 16287, Application US/09949016
;; Patent No. 6812339
;; GENERAL INFORMATION:
;; APPLICANT: VENTUR, J. Craig et al.
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;; FILE REFERENCE: CL001307
;; CURRENT APPLICATION NUMBER: US/09/949,016
;; CURRENT FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/241,755
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FASTSEQ for Windows Version 4.0
;; SEQ ID NO 16287
;; LENGTH: 60489
;; TYPE: DNA
;; ORGANISM: Human
US-09-949-016-16287

Query Match
Best Local Similarity 69.2%; Score 54.2; DB 4; Length 60489;
Pred. No. 0.0006;
Matches 74; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 490 CTCACCTCAGTTTGTGATCCGTTAAATGAGCAAAATTCGAACTTCACTGTCGT 549
DB 12018 CTGTAAGCTGTTTCTTCCATGTTGTAACAAGGATTAATAAAGCTTCACTGTCGT 11959
QY 550 TGAAGAGATTAAATGAACAATGCTTTGTAAGCTCTTGGAGAGGG 596
DB 11958 TGTGAGATGAATCTATTATACATGTAAGCACTTTGAACAATGG 11912

RESULT 4
US-09-949-016-14157/c
;; Sequence 14157, Application US/09949016
;; Patent No. 6812339
;; GENERAL INFORMATION:
;; APPLICANT: VENTUR, J. Craig et al.
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;; FILE REFERENCE: CL001307

;; CURRENT APPLICATION NUMBER: US/09/949,016
;; CURRENT FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/241,755
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FASTSEQ for Windows Version 4.0
;; SEQ ID NO 14157
;; LENGTH: 455726
;; TYPE: DNA
;; ORGANISM: Human
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)...(455726)
;; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14157

Query Match
Best Local Similarity 61.8%; Score 52.8; DB 4; Length 455726;
Pred. No. 0.0039;
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 474 TGACCTACACATGTGACTTCACTCAGTTTGTGATCCGTTAAATGAGCAAAATTCGAGC 533
DB 38822 TGGCTGGGAGAGTAATCTTCACTGATGTTCTCATCTGAACATGGTACGTGGCC 38763
QY 534 TACTTCAAGTGTGTTGAGAGATTAAATGAACAATCTGTGTAAGCTCTTGGAGCA 593
DB 38762 TGATTCAACAAGGTGTTGTAAGATTAACTGACATGCTGTGAAGAAGCACTTACGAGA 38703
QY 594 GCGAGCTCGGAAGCA 609
DB 38702 TGCTTAGTCCAAGCA 38687

RESULT 5
US-09-949-016-11940/c
;; Sequence 11940, Application US/09949016
;; Patent No. 6812339
;; GENERAL INFORMATION:
;; APPLICANT: VENTUR, J. Craig et al.
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;; FILE REFERENCE: CL001307
;; CURRENT APPLICATION NUMBER: US/09/949,016
;; CURRENT FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/241,755
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FASTSEQ for Windows Version 4.0
;; SEQ ID NO 11940
;; LENGTH: 48115
;; TYPE: DNA
;; ORGANISM: Human
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)...(48115)
;; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11940

Query Match
Best Local Similarity 61.8%; Score 52.8; DB 4; Length 48115;
Pred. No. 0.004;
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 474 TGACCTACACATGTGACTTCACTCAGTTTGTGATCCGTTAAATGAGCAAAATTCGAGC 533
DB 16211 TGGCTGGGAGAGTAATCTTCACTGATGTTCTCATCTGAACATGGTACGTGGCC 16152

LENGTH: 132871
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(132871)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13863

Query Match
Best Local Similarity 45.9%; Score 49.6; DB 4; Length 132871;
Pred. No. 0.017;
Matches 204; Conservative 0; Mismatches 239; Indels 1; Gaps 1;

QY 8 GCGGCTCTGCGGCGCTGATACACAGCAGCCTGATGATGAGCAACAGAGAGCCCTCTCC 67
DB 1734 GAGGCGGGGGGCGCCCGGGGCGCTCCCGCCCGCGCGCGCGCGCGCGCTTAAAGC 1793
QY 68 ATGTTTAGGACCTCTGAGGCGCTGAGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCG 127
DB 1794 CGGGCGGGGCG 1853
QY 128 ATCCGCGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 187
DB 1854 GCGGCGGGGCG 1913
QY 188 TGGGTGCAAGTCTGAGGCTTCTGAGTTCTGAGTTCTGAGTTCTGAGTTCTGAGTT 247
DB 1914 CGGCGGGGCG 1973
QY 248 GCGGCTCTGCGGCTGAGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 307
DB 1974 G-CGGGCTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2032
QY 308 GGTCCGCGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 367
DB 2033 AGCGGCGGTTGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2092
QY 368 CCCCCGCGCTCAGCAGCGGCTCAGCAGCTGATGAGGAGGAGGAGGAGGAGGAGG 427
DB 2093 TCTTCCCG 2152
QY 428 TGAGATCCCGATGCTTACGAGCC 451
DB 2153 GAGCG 2176

RESULT 10
US-09-949-016-12210
Sequence 12210, Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12210
LENGTH: 161607
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(161607)
OTHER INFORMATION: n = A,T,C or G

US-09-949-016-12210

Query Match
Best Local Similarity 45.9%; Score 49.6; DB 4; Length 161607;
Pred. No. 0.018;
Matches 204; Conservative 0; Mismatches 239; Indels 1; Gaps 1;

QY 8 GCGGCTCTGCGGCGCTGATACACAGCAGCCTGATGATGAGCAACAGAGAGCCCTCTCC 67
DB 1734 GAGGCGGGGGGCGCCCGGGGCGCTCCCGCCCGCGCGCGCGCGCGCGCTTAAAGC 1793
QY 68 ATGTTTAGGACCTCTGAGGCGCTGAGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCG 127
DB 1794 CGGGCGGGGCG 1853
QY 128 ATCCGCGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 187
DB 1854 GCGGCGGGGCG 1913
QY 188 TGGGTGCAAGTCTGAGGCTTCTGAGTTCTGAGTTCTGAGTTCTGAGTTCTGAGTT 247
DB 1914 CGGCGGGGCG 1973
QY 248 GCGGCTCTGCGGCTGAGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 307
DB 1974 G-CGGGCTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2032
QY 308 GGTCCGCGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 367
DB 2033 AGCGGCGGTTGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2092
QY 368 CCCCCGCGCTCAGCAGCGGCTCAGCAGCTGATGAGGAGGAGGAGGAGGAGGAGG 427
DB 2093 TCTTCCCG 2152
QY 428 TGAGATCCCGATGCTTACGAGCC 451
DB 2153 GAGCG 2176

RESULT 11
US-09-949-016-50874
Sequence 50874, Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 50874
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-50874

Query Match
Best Local Similarity 64.0%; Score 48.4; DB 4; Length 601;
Pred. No. 0.0027;
Matches 73; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 477 CCTACATGTGACTTCACTCAGTTTGTGATCCGTAATAAGACAAATTCAGAGTAC 536
DB 178 CCAAGACAGAGTAACTTCACTCAGTTTGTGATCCGTAATAAGACAAATTCAGAGTAC 237
QY 537 TTCACTGCTGTTGAGAGATTAAATGAACAATGCTTGAAGCTTTGCA 590

Db 238 TTCAATGCTGTTGTGAGCATTAACAAGAACCTACATTAAGATTACCA 291

RESULT 12

US-09-949-016-13214
; Sequence 13214, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13214
; LENGTH: 113701
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13214

Query Match 2.7%; Score 48.4; DB 4; Length 113701;
Best Local Similarity 64.0%; Pred. No. 0.034;
Matches 73; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 477 CTTACACATGTCACCTTCACTTCACTTGTGATCCGTAATAATGCAAAATTCGAAGCTAC 536
Db 102206 CCAACAGCAAGTGAATCACTTCACTTCACTTGTGATCCGTAATAATGCAAAATTCGAAGCTAC 102265

Qy 537 TTCACAGCTGCTGTGAGAGATTAAATGAACAATGCTTGAACCTCTTGGCA 590
Db 102266 TTCAATGCTGTTGTGAGCATTAACAAGAACCTACATTAAGATTACCA 102319

RESULT 13

US-09-949-016-11815/c
; Sequence 11815, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11815
; LENGTH: 26845
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11815

Query Match 2.7%; Score 48.2; DB 4; Length 26845;
Best Local Similarity 56.7%; Pred. No. 0.019;
Matches 89; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 490 CTTACCTCAGTTTGTGATCGTAATAATGCAAAATTCGAAGCTACTTCAAGTGTGT 549

Db 13701 CTGACCTTTTGTTCCTCATTTGTAATAAGGTGAGTATGTCTGTACAGTGTGT 13642

Qy 550 TGAGAGATTAAATGAACAATGCTTGTAAAGCTTTTGCAGAGAGACCTCGAGACA 609

Db 13641 TGTGGAATTAAATGATGAATAAGGCTATTAATCTTGTGCAAGACATTAAATGTG 13582

Qy 610 GGGCTGCGCGGACAGACACCTGCTGTCAACGAGG 646

Db 13581 GGGTAGAGAGAGTAAAGTAAACAATTGCGACGAGTG 13545

RESULT 14

US-09-949-016-15424/c
; Sequence 15424, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15424
; LENGTH: 27132
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15424

Query Match 2.7%; Score 48.2; DB 4; Length 27132;
Best Local Similarity 56.7%; Pred. No. 0.019;
Matches 89; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 490 CTTACCTCAGTTTGTGATCGTAATAATGCAAAATTCGAAGCTACTTCAAGTGTGT 549

Db 13988 CTGACCTTTTGTTCCTCATTTGTAATAAGGTGAGTATGTCTGTACAGTGTGT 13929

Qy 550 TGAGAGATTAAATGAACAATGCTTGTAAAGCTTTTGCAGAGAGACCTCGAGACA 609

Db 13928 TGTGGAATTAAATGATGAATAAGGCTATTAATCTTGTGCAAGACATTAAATGTG 13869

Qy 610 GGGCTGCGCGGACAGACACCTGCTGTCAACGAGG 646

Db 13868 GGGTAGAGAGAGTAAAGTAAACAATTGCGACGAGTG 13832

RESULT 15

US-09-949-016-13922
; Sequence 13922, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13922


```

QY 317 CCCGAGCATTCGGGGCTGCAGATTGACGGGGATCCCGGATGACCCGGCGCCCCCGCC 376
Db 130 CCCGAGGATTCGGGGCTGCAGATTGACGGGGATCCGGAGTGCACCGGCGCCCCCGGCG 189
QY 377 CCTGACCGACGGGTTCAGACCTGGTGGGAAGAGGTGGGGAGCGGCTCCTGAGATCC 436
Db 190 CCTGACCGAGGGGTTCAGACCTGGTGGGAAGAGGTGGGGAGCGGCTCCTGAGATCC 249
QY 437 CGATGCTTACGAGCCAGATGCTCACTTTATAGGTGTGACCTTACATGTGACTTCAAC 496
Db 250 CGATGCTTACGAGCCAGATGCTCACTTTATAGGTGTGACCTTACATGTGACTTCAAC 309
QY 497 TCAGTTTGTGATCCGTAAATGGAACAATCGAAGCTACTTCAAGGCTGTGAGAGG 556
Db 310 TCAGTTTGTGATCCGTAAATGGAACAATCGAAGCTACTTCAAGGCTGTGAGAGG 369
QY 557 ATTAATGAAAACAATGCTTTGTAAGCTTTTGACAGAGGAGCCTCGAAGCAGGGCTG 616
Db 370 ATTAATGAAAACAATGCTTTGTAAGCTTTTGACAGAGGAGCCTCGAAGCAGGGCTG 429
QY 617 GCCGGCAGAGACACCTGCTGTACACAGGACCAAGGAGCATGAAAGATCCCGTGGAG 676
Db 430 GCCGGCAGAGACACCTGCTGTACACAGGACCAAGGAGCATGAAAGATCCCGTGGAG 489
QY 677 CTGAGCCGTCAAGTGGAGTGCAGACCTCGGCTTGCAGACCGGCTGCGAGGCTTACCGG 736
Db 490 CTGAGCCGTCAAGTGGAGTGCAGACCTCGGCTTGCAGACCGGCTGCGAGGCTTACCGG 549
QY 737 GTCAAGGCCAGAGCATATATGTGATGAGACTCTGTTTGGCAGCCGACAGGACCCGG 796
Db 550 GTCAAGGCCAGAGCATATATGTGATGAGACTCTGTTTGGCAGCCGACAGGACCCGG 609
QY 797 CTTACCCCAACCGGACTTTCGATCCGCCCTGGGTGGAAGAGGCTTAAACAACACAGGGCTG 856
Db 610 CTTACCCCAACCGGACTTTCGATCCGCCCTGGGTGGAAGAGGCTTAAACAACAGGGCTG 669
QY 857 GGCAAGAGGAGCATGAGAGGCTTGGGGGCAAAAGGGAGGCTGTGAGACCAACCCCTCAAG 916
Db 670 GGCAAGAGGAGCATGAGAGGCTTGGGGGCAAAAGGGAGGCTGTGAGACCAACCCCTCAAG 729
QY 917 GGCAAGACCCCAACCTTCAACAAGAGAGAAACAATTAACAAGCCATCAAGCACACC 976
Db 730 GGCAAGACCCCAACCTTCAACAAGAGAGAAACAATTAACAAGCCATCAAGCACACC 789
QY 977 CCGCTTACTGTGATGAGTGGCTGGTGGCTCCGATCTGAAGGCGCACCTTCGGGGCC 1036
Db 790 CCGCTTACTGTGATGAGTGGCTGGTGGCTCCGATCTGAAGGCGCACCTTCGGGGCC 849
QY 1037 CCGCGATGGAGAGGGGGAGTGCAGCAAGCTCGTGCTCTTGTGAGACGACCAACT 1096
Db 850 CCGCGATGGAGAGGGGGAGTGCAGCAAGCTCGTGCTCTTGTGAGACGACCAACT 909
QY 1097 ACCGCCAGGGGTAGCCACTGGCCCGGCCCAAGGAGCCACTGCGAGCCATTCACCCA 1156
Db 910 ACCGCCAGGGGTAGCCACTGGCCCGGCCCAAGGAGCCACTGCGAGCCATTCACCCA 969
QY 1157 GCTGATCCCTTCAAGAGAGAGCGGGGCAAGCGGCAAGCTCCAGAAAGTATCTAAGGGT 1216
Db 970 GCTGATCCCTTCAAGAGAGAGCGGGGCAAGCGGCAAGCTCCAGAAAGTATCTAAGGGT 1029
QY 1217 GGGTTACACTTTCACGCGCCCTGMAAGCGGGGACTTTCCTCATTCCTCAACCACTGAAT 1276
Db 1030 GGGTTACACTTTCACGCGCCCTGMAAGCGGGGACTTTCCTCATTCCTCAACCACTGAAT 1089
QY 1277 GTCCCCAGACCTGTCATTCAGCCACAGTGCCTCCCAACAATGGGGCTCAGATTC 1336
Db 1090 GTCCCCAGACCTGTCATTCAGCCACAGTGCCTCCCAACAATGGGGCTCAGATTC 1149
QY 1337 AGGCTTTCACAGTTCAGGGGTGACCTTCCGAGAGCCCTCTGTGACTTCAAGGGCTCGTCA 1396
Db 1150 AGGCTTTCACAGTTCAGGGGTGACCTTCCGAGAGCCCTCTGTGACTTCAAGGGCTCGTCA 1209

```

```

QY 1397 GTTAGCATTTCAAGCCATCTAACCCACGACGAGTGGGGGCCAACCCAGAAACCAAGCCC 1456
Db 1210 GTTAGCATTTCAAGCCATCTAACCCACGACGAGTGGGGGCCAACCCAGAAACCAAGCCC 1269
QY 1457 CTTGGAATATGATATCTTTTCATCAGGGTGGCTTAATGGGGCCACGCGCAGGATATGGCC 1516
Db 1270 CTTGGAATATGATATCTTTTCATCAGGGTGGCTTAATGGGGCCACGCGCAGGATATGGCC 1329
QY 1517 CTTTCCAGGGGTGAGAGGACATTCATCACCAGGGAAACCCAGATATTAAGAACCCCT 1576
Db 1330 CTTTCCAGGGGTGAGAGGACATTCATCACCAGGGAAACCCAGATATTAAGAACCCCT 1389
QY 1577 GTGGGGGACAGACGACATAGCAGGGGTGGGAGGAGGCTCCCTTATCTGACATCTTA 1636
Db 1390 GTGGGGGACAGACGACATAGCAGGGGTGGGAGGAGGCTCCCTTATCTGACATCTTA 1449
QY 1637 GTGATTTCTGCTTTTCTCCGATTTGCGGATTTGGGGGCCACTTAAGATGCTCTC 1696
Db 1450 GTGATTTCTGCTTTTCTCCGATTTGCGGATTTGGGGGCCACTTAAGATGCTCTC 1509
QY 1697 TCCAGCCCTGTTCACACATCTTCAAATTAATGTCACACCCAGGGGCTTGGCACTTCCA 1756
Db 1510 TCCAGCCCTGTTCACACATCTTCAAATTAATGTCACACCCAGGGGCTTGGCACTTCCA 1569
QY 1757 CATCATCATTTGCTTGTGCTGCCAAGTGCAGATTAACGGGCGTGAATGCAACTGG 1811
Db 1570 CATCATCATTTGCTTGTGCTGCCAAGTGCAGATTAACGGGCGTGAATGCAACTGG 1624

```

```

RESULT 2
US-10-452-858C-78
; Sequence 78; Application US/10452858C
; Publication No. US20040086945A1
; GENERAL INFORMATION:
; APPLICANT: Streekrishna, Kocikanyadanam
; APPLICANT: Gerwe, Gina S.
; APPLICANT: Toerner, Daniel R.
; TITLE OF INVENTION: HAIRLESS PROTEIN-INTERACTING PARTNER COMPLEXES AND METHODS THEREOF
; FILE REFERENCE: 8956P
; CURRENT APPLICATION NUMBER: US/10/452, 858C
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78
; LENGTH: 709
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21)..(707)
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (670)..(695)
; OTHER INFORMATION: n represents a, c t or g
US-10-452-858C-78

```

```

Query Match 38.0%; Score 688.2; DB 17; Length 709;
Best Local Similarity 99.1%; Pred. No. 5,8e-200;
Matches 690; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 682 CGTCACTGGAGATGACAGCCCTCGGCTTTCAGACACGCTCCGAGGTGGCTACCGGGTCAA 741
Db 14 CGTCACTGGAGATGACAGCCCTCGGCTTTCAGACACGCTCCGAGGTGGCTACCGGGTCAA 73
QY 742 GGCAGAGAGCTCATATGTGATGAGACTCTGTTTGGCAGCCGACAGGAGGACCCGGGCTAC 801
Db 74 GGCAGAGAGCTCATATGTGATGAGACTCTGTTTGGCAGCCGACAGGAGGACCCGGGCTAC 133
QY 802 CCACACCGACTTCGATCCGCCCTGGGTGGAAGGTTAACAAACCAAGAGGGGTGGGCAA 861
Db 134 CCACACCGACTTCGATCCGCCCTGGGTGGAAGGTTAACAAACCAAGAGGGGTGGGCAA 193

```

QY 862 GGAGGATCGAAGGCTTGGGGGCAAGGGAGCTGTAGACACCCCTCAAGGGGAG 921
DB 194 GGAGGATCGAAGGCTTGGGGGCAAGGGAGCTGTAGACACCCCTCAAGGGGAG 253
QY 922 CACCCCTTCAACCAAGGAGAAACAATACAGACCCATCAGCCACACCCCTG 981
DB 254 CACCCCTTCAACCAAGGAGAAACAATACAGACCCATCAGCCACACCCCTG 313
QY 982 TTACTGTAGTGGCTTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 1041
DB 314 TTACTGTAGTGGCTTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 373
QY 1042 GATGCGAAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1101
DB 374 GATGCGAAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 433
QY 1102 CAGGGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1161
DB 434 CAGGGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 493
QY 1162 TCCCTCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1221
DB 494 TCCCTCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 553
QY 1222 ACACCTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1281
DB 554 ACACCTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 613
QY 1282 CAGCACTGATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1341
DB 614 CAGCACTGATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 673
QY 1342 TTCACGTGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1377
DB 674 TTCACGTGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 709

RESULT 3

US-09-918-995-1667
; Sequence 1667, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1667
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(477)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-1667

Query Match

24.1%; Score 436.6; DB 10; Length 477;
Best Local Similarity: 99.1%; Pred. No. 5.9e-123;
Matches 439; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 986 TGTGATGATGCTGCTTTGGCTCCGATCTGAAGGCGCAGCTTGGGGGCGCGGAG 1045
DB 35 TGTGATGATGCTGCTTTGGCTCCGATCTGAAGGCGCAGCTTGGGGGCGCGGAG 94
QY 1046 GCGAAGGGGAGTGCAGCAAGCTCCGTCTCTTGTGAGCGCCACCACTTACCCCA 1105
DB 95 GCGAAGGGGAGTGCAGCAAGCTCCGTCTCTTGTGAGCGCCACCACTTACCCCA 154

QY 1106 GGTAGCACTGCGCCCGCCGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1165
DB 155 GGTAGCACTGCGCCCGCCGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 214
QY 1166 TCCAGAGAGAGGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1225
DB 215 TCCAGAGAGAGGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 274
QY 1226 TTTTCAAGGCGGCTTGAAGCGGGGAGTTCCTTCCATCCACCTGATATCCGAC 1285
DB 275 TTTTCAAGGCGGCTTGAAGCGGGGAGTTCCTTCCATCCACCTGATATCCGAC 334
QY 1286 ACTGTCATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1345
DB 335 ACTGTCATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 394
QY 1346 AGTCAGGGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1405
DB 395 AGTCAGGGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 454
QY 1406 TCAGTCCATCTTACCCCAAGAG 1428
DB 455 TCAGTCCATCTTACCCCAAGAG 477

RESULT 4

US-09-918-995-31166
; Sequence 31166, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31166
; LENGTH: 484
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(484)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-31166

Query Match 22.7%; Score 410.6; DB 10; Length 484;
Best Local Similarity: 93.9%; Pred. No. 5.4e-115;
Matches 419; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 426 CTTGAGATCCCGATGCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 485
DB 39 CTTGAGATCCCGATGCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 98
QY 486 GTGACTTCACTGATTTTGTGATCCGTAATATGAGCAATTTGAAAGCTTTCACAG 545
DB 99 GTGACTTCACTGATTTTGTGATCCGTAATATGAGCAATTTGAAAGCTTTCACAG 158
QY 546 CTTTGAAGAGATTAATGAACAATGCTTTGAAGCTTTTGAAGAGGAGGAGGAG 605
DB 159 CTTTGAAGAGATTAATGAACAATGCTTTGAAGCTTTTGAAGAGGAGGAGGAG 218
QY 606 AGCAGGGCTTGGCGGAGAGACACCTGCTGTACCAAGGAGACCAAGGAGGAGGAG 665
DB 219 AGCAGGGCTTGGCGGAGAGACACCTGCTGTACCAAGGAGACCAAGGAGGAG 278
QY 666 CCCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 725

Db 279 CCCCCGTGAGCTGGCCCTGTCAGTGAGATGACACCTTGCGCTTCAACACCGCTGCCGAG 338
QY 726 GTGGCTACCGGGGTCAAGGCCGACGTGATATGTGATGATGATCTCTGTTGGCAGCCGAG 785
Db 339 GTGGCTACCGGGGTCAAGGCCGACGTGATATGTGATGATGATCTCTGTTGGCAGCCGAG 398
QY 786 CAGGACACCGGGCTTACCCCAACCGGACTTGCATCCGCTGGGTGAGAGAGGCTAAGAGAA 845
Db 399 CAGGACACCGGGCTTACCCCAACCGGACTTGCATCCGCTGGGTGAGAGAGGCTAAGAGAA 458
QY 846 CCAAGAGCGGTGGGCAAGAGGCGATCG 871
Db 459 CCAGAGGCGGTGGGCAAGAGGCGATCG 484

RESULT 5
US-09-864-408A-8425
; Sequence 8425, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 8425
; LENGTH: 292
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-864-408A-8425

Query Match 7.2%; Score 130.8; DB 11; Length 292;
Best Local Similarity 72.8%; Pred. No. 2.3e-29;
Matches 182; Conservative 0; Mismatches 67; Indels 1; Gaps 1;

QY 639 CACGAGGAGCCACGACGATGAAGACCCCGTGG-AGCTGGCCGTGATGGGATGACAG 697
Db 43 CACGAGGAGCCACGACGATGAAGACCCCGTGGCAGCCGCGCCACGAGCATGATAA 102
QY 698 ACCCTGGGCTTACGACACCGCTGCGAGTGGCTACACCGGGTCAAGGCGAGAGCTGAT 757
Db 103 GCATTCACCTTCAGCAGCCGAGTCCACAGCTACAGGGTCAAGGCGAGGATCTTAT 162
QY 758 GTGATGAGACTGTTTGGCAGCCGACGAGGCAACCGGCTTACCCACCGGACTTGAT 817
Db 163 GTGATGAGACTGTTTGGCAGCCGACGAGGCAACCGGCTTACCCACCGGACTTGAT 222
QY 818 CCGGCTGGGTGAGAGGCTTAACAGAACGAGGCGCTGGGCAAGAGGATGGAAGGC 877
Db 223 CCACCTGGGTGAGAGGCTTAACAGAACGAGGCGCTGGGCAAGAGGATGGAAGGC 282
QY 878 TTGGGGGCA 887
Db 283 TCTCTGGCA 292

RESULT 6
US-09-908-975-4892
; Sequence 4892, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Ilat
; APPLICANT: PALGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLIC
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME

FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4892
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-4892

Query Match 3.3%; Score 60; DB 10; Length 60;
Best Local Similarity 100.0%; Pred. No. 7e-08;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 TGGTGGCTGCTCCCTGGTGTCTGAGTGCAGAAAGTGGGTTCTGGGTTCTGGATTC 223
Db 1 TGGTGGCTGCTCCCTGGTGTCTGAGTGCAGAAAGTGGGTTCTGGGTTCTGGATTC 60

RESULT 7
US-10-027-632-288030
; Sequence 288030, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 288030
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-288030

Query Match 2.9%; Score 52.8; DB 13; Length 479;
Best Local Similarity 61.8%; Pred. No. 2e-05;
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 474 TGACCTACAGATGTGACTTACCTCAGTTTGTGATCCGTAATAATGACAAATTCGAGC 533
Db 89 TGGCCGTGGGAGGTAACTTACCTCAGTTTGTGATCCGTAATAATGACAAATTCGAGC 148
QY 534 TACTTACAGTGTGTTGAGAGATTAATGAACAGTCTGTAAAGCTCTTTGCAGGA 593
Db 149 TGATTCACAAGGTGTTGTAAGATTAAGTACATGACGTGTGAAGAGCACTTAAGAGA 208
QY 594 GGGAGCCTCGGAAGCA 609
Db 209 TGCTTAGTCCAAAGCA 224

```

?      LENGTH: 493599
?      TYPE: DNA
?      ORGANISM: Homo sapiens
?      FEATURE:
?      NAME/KEY: misc_feature
?      LOCATION: (1)...(493599)
?      OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
?      IS-10-719-993-6787

```

	Query Match	2.8%	Score 51.6	DB 17	Length 680
	Best Local Similarity	67.9%	Pred. No. 5e-05		
	Matches	72	Conservative	0	Mismatches 34
				Indels	0
				Gaps	0
Oy	490	CTTCACCTAGATTGTAATCGGTAAATGGACAATTCGAAAGTACTCTCAAGTGCCT	54		
Db	35	CCTTACCTCAATCTCTTATCTGTGAATGGGAATTAATTAATCTCACTCACAAGATTAT	94		
Oy	550	TGAGAGATTTAATGAACAATGCTTTGTAAAGCTCTTGCAGAGG	595		
Db	95	TAAAGAAATTAATGAGTAATCCATGGAAAGCTAGTAGCGAGATG	140		

RESULT 11
US-10-242-355-1102
; Sequence 1102, Application US/10242355

```
; Publication No. US20030235831A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC003C1
; CURRENT APPLICATION NUMBER: US/10/242,355
; PRIOR FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,897
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1102
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-242-355-1102

Query Match      2.8%; Score 51.6; DB 17; Length 680;
Best Local Similarity 67.9%; Pred. No. 5.3e-05;
Matches 72; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY      490 CTTCACTCAGTTTGGATCCGTAATAATGCAAAATTCGAAGCTTCTACAGTGTGT 549
Db      35 CCTTACCTCAATCTCCTTATCTGTGAATGCGAATATAAATCTACCTACAGAAATTAT 94

QY      550 TGAGAGATTAAATGAACAATGCTTTAAAGCTCTTGAGAGAG 595
Db      95 TAACAGATTAAATGAGATTATTCATGGAAGCTAGTAGCAGCATG 140

RESULT 12
US-10-242-355-240
; Sequence 240, Application US/10242355
; Publication No. US20030235831A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC003C1
; CURRENT APPLICATION NUMBER: US/10/242,355
; PRIOR FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,897
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
```

```
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 240
; LENGTH: 343
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-242-355-240

Query Match      2.8%; Score 50.8; DB 17; Length 343;
Best Local Similarity 66.0%; Pred. No. 7.6e-05;
Matches 70; Conservative 2; Mismatches 34; Indels 0; Gaps 0;

QY      490 CTTCACTCAGTTTGGATCCGTAATAATGCAAAATTCGAAGCTTCTACAGTGTGT 549
Db      35 CCTTACCTCAATCTCCTTATCTGTGAATGCGAATATAAATCTACCTACAGAAATTAT 94

QY      550 TGAGAGATTAAATGAACAATGCTTTAAAGCTCTTGAGAGAG 595
Db      95 TAACAGATTAAATGAGATTATTCATGGAAGCTAGTAGCAGCATG 140

RESULT 13
US-10-027-632-102308/c
; Sequence 102308, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 102308
; LENGTH: 741
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-102308

Query Match      2.8%; Score 50.4; DB 13; Length 741;
Best Local Similarity 60.0%; Pred. No. 0.00013;
Matches 84; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY      487 TGACTTCACCTCAGTTTGGATCCGTAATAATGCAAAATTCGAAGCTTCTACAGTGT 546
Db      444 TACCTAAGACACAGTTTCTTTATCTGTAAATAGTACATGATGACATCTTGAAAGGCG 385

QY      547 TGTGAGAGATTAAATGAACAATGCTTTAAAGCTCTTTGAGAGAGGAGCTTGGA 606
Db      384 TACTGTGAGAGATTATATGAATAATGAACGCAATATTAACATTAAGAGTTCTGAGCTCAAAA 325
```

QY 607 GCAGGCGCTGGCCGCGAG 626
 DB 324 AGGCACTGTATGATGAG 305

RESULT 14

US-10-027-632-102308/c
 ; Sequence 102308, Application US/10027632
 ; Publication No. US20030204075A9
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, David G.
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
 ; FILE REFERENCE: 108827.129
 ; CURRENT APPLICATION NUMBER: US/10/027,632
 ; CURRENT FILING DATE: 2002-04-30
 ; PRIOR APPLICATION NUMBER: US 60/218,006
 ; PRIOR FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: US 60/198,676
 ; PRIOR FILING DATE: 2000-04-20
 ; PRIOR APPLICATION NUMBER: US 60/193,483
 ; PRIOR FILING DATE: 2000-03-29
 ; PRIOR APPLICATION NUMBER: US 60/185,218
 ; PRIOR FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/167,363
 ; PRIOR FILING DATE: 1999-11-23
 ; PRIOR APPLICATION NUMBER: US 60/156,358
 ; PRIOR FILING DATE: 1999-09-28
 ; PRIOR APPLICATION NUMBER: US 60/146,002
 ; PRIOR FILING DATE: 1999-08-09
 ; NUMBER OF SEQ ID NOS: 325720
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 102308
 ; LENGTH: 741
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-10-027-632-102308

Query Match 2.8%; Score 50.4; DB 17; Length 741;
 Best Local Similarity 60.0%; Pred. No. 0.00013;
 Matches 84; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 487 TGACTTCACCTGATTTGTGATCCGTAAATGACAAATTCGAGCTTACACAGTGC 546
 DB 444 TCACCTAAGACACAGTTCTTTATCTGTAAATAGTACATGATGATCTTGAAGAGGC 385
 QY 547 TGTGAGAGGATTAATGAACAATGCTTGTAAAGCTTTTCAGAGGAGCCTTGGA 606
 DB 384 TACTGTGAGATTATATGAAATGAAACGCAATGATTAAGGTTCTGAGACTCAAA 325
 QY 607 GCAGGCGCTGGCCGCGAGAG 626
 DB 324 AGGCACTGTATGATGAG 305

RESULT 15

US-10-292-798-1243
 ; Sequence 1243, Application US/10292798
 ; Publication No. US20030235833A1
 ; GENERAL INFORMATION:
 ; APPLICANT: SUMA, MAKIKO
 ; APPLICANT: ASAI, KIYOSHI
 ; APPLICANT: AKIYAMA, YUTAKA
 ; APPLICANT: ABRUATANI, HIROYUKI
 ; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
 ; FILE REFERENCE: 084335/166
 ; CURRENT APPLICATION NUMBER: US/10/292,798
 ; CURRENT FILING DATE: 2002-11-13
 ; PRIOR APPLICATION NUMBER: 10/017,161
 ; PRIOR FILING DATE: 2001-12-18
 ; PRIOR APPLICATION NUMBER: JP 2001-246789
 ; PRIOR FILING DATE: 2001-06-18
 ; NUMBER OF SEQ ID NOS: 2070

SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1243
 ; LENGTH: 43981
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; LOCATION: source
 ; FEATURE:
 ; LOCATION: (1)..(43981)
 ; NAME/KEY: CDS
 ; LOCATION: (201)..(470)
 ; FEATURE:
 ; LOCATION: (12636)..(12748)
 ; NAME/KEY: CDS
 ; LOCATION: (24946)..(25088)
 ; FEATURE:
 ; LOCATION: (32801)..(32997)
 ; NAME/KEY: CDS
 ; LOCATION: (41097)..(41379)
 ; FEATURE:
 ; LOCATION: (2811)..(2910)
 ; NAME/KEY: modified base
 ; LOCATION: (6914)..(6914)
 ; OTHER INFORMATION: a, t, c, g, unknown or other
 ; FEATURE:
 ; LOCATION: (6923)..(6923)
 ; NAME/KEY: modified base
 ; LOCATION: (6956)
 ; OTHER INFORMATION: a, t, c, g, unknown or other
 ; FEATURE:
 ; LOCATION: (6932)..(6932)
 ; NAME/KEY: modified base
 ; LOCATION: (6956)
 ; OTHER INFORMATION: a, t, c, g, unknown or other
 ; FEATURE:
 ; LOCATION: (6959)..(6960)
 ; NAME/KEY: modified base
 ; LOCATION: (6964)
 ; OTHER INFORMATION: a, t, c, g, unknown or other
 ; FEATURE:
 ; LOCATION: (6964)
 ; NAME/KEY: modified base
 ; LOCATION: (7905)
 ; OTHER INFORMATION: a, t, c, g, unknown or other
 ; US-10-292-798-1243

Query Match 2.8%; Score 50.4; DB 17; Length 43981;
 Best Local Similarity 71.7%; Pred. No. 0.00042;
 Matches 66; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 495 CCTGAGTTTGTGATCCGTAAATGACAAATTCGAGCTTACAGTGTGTGGA 554
 DB 18915 CCTGAGTTTGTGATCCGTAAATGACAAATTCGAGCTTACAGTGTGTGGA 18974
 QY 555 GGATTAATGAACAATGCTTGTAAAGCTTT 586
 DB 18975 GGATTAATGAACAATGCTTGTAAAGCTTT 19006

Mon Mar 28 09:43:46 2005

us-10-031-589-3.rmpb

Page 8

Search completed: March 28, 2005, 08:50:19
Job time : 2119 secs

Result	Query			ID		Description
No.	Score	Match	Length	DB	ID	
1	1764.4	97.4	1766	3	CR595506	CR595506 full-1-1eng
2	1764.4	97.4	1936	3	CR608011	CR608011 full-1-1eng
3	1659	91.6	1839	3	CR623111	CR623111 full-1-1eng
4	1631	90.1	1632	3	CR621376	CR621376 full-1-1eng
5	1614	89.1	1614	3	CR621332	CR621332 full-1-1eng
6	1600	88.3	1780	3	CR600183	CR600183 full-1-1eng
7	1596	88.1	1596	3	CR597629	CR597629 full-1-1eng
8	1557	86.0	1575	3	CR606890	CR606890 full-1-1eng
9	1541	85.1	1560	3	CR625812	CR625812 full-1-1eng
10	1346	74.3	1488	3	CR611377	CR611377 full-1-1eng
11	1325	73.2	1467	3	CR622791	CR622791 full-1-1eng
12	1323	73.1	1465	3	CR605763	CR605763 full-1-1eng
13	1045	57.7	1045	3	CR594289	CR594289 full-1-1eng
14	1012.2	55.9	1098	5	BX358473	BX358473 full-1-1eng
15	976.8	53.9	1038	5	BX339727	BX339727 full-1-1eng
16	971.2	53.6	1008	5	BX339726	BX339726 full-1-1eng
17	932.4	51.5	1101	1	AL523765	AL523765 full-1-1eng
18	930.4	51.4	1113	5	BX403928	BX403928 full-1-1eng
19	924.4	51.0	1063	1	AL560267	AL560267 full-1-1eng
20	909.2	50.2	951	5	BX354462	BX354462 full-1-1eng
21	898.6	49.6	998	5	BX345478	BX345478 full-1-1eng
22	892	49.3	892	3	CR596541	CR596541 full-1-1eng
23	889.2	49.1	1006	1	AL582549	AL582549 full-1-1eng
24	884.2	48.8	1047	5	BM903591	BM903591 full-1-1eng

C	25	880.6	48.6	991	1	AL582286	AL582286
	26	880.6	48.6	1009	1	AL517881	AL517881
	27	876.8	48.4	1046	5	BX384471	BX384471
	28	876.6	48.4	1029	5	BX376800	BX376800
	29	862.2	47.4	1032	5	BX332260	BX332260
	30	860.2	47.5	1017	5	BM924108	BM924108
	31	857.6	47.4	972	1	AL561282	AL561282
	32	852	47.0	1063	1	AL560832	AL560832
	33	849	46.9	983	5	BX353787	BX353787
	34	845.8	46.7	926	5	BX394189	BX394189
C	35	835.4	45.1	872	5	BX346001	BX346001
	36	830.2	45.18	866	5	BX352694	BX352694
	37	826.8	45.17	950	5	BQ943060	BQ943060
C	38	826.6	45.6	1047	1	AL581746	AL581746
	39	823.4	45.5	856	5	BX366971	BX366971
	40	815	45.0	892	1	AL516514	AL516514
	41	809.6	44.7	879	5	BU526901	BU526901
	42	806	44.5	903	1	AL555180	AL555180
	43	803.4	44.4	907	5	BQ962573	BQ962573
	44	791.6	43.7	1011	5	BX366455	BX366455
	45	785.8	43.4	1038	5	BX356915	BX356915

ALIGNMENTS

RESULT 1
CR595506

DEFINITION full-length cDNA clone CS0DG007YI18 of B cells (Ramos cell line) of *Homo sapiens* (human).

ACCESSION	CR595506
VERSION	CR595506.1
	GI:50476313

KEYWORDS	HTC; CNSLT_cDNA.
SOURCE	Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; M

REFERENCE
1 (bases 1 to 1766)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization

JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifestech.com URL :

<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600
Faraday Avenue

REFERENCE	2 (bases 1 to 1766)
AUTHORS	Genoscope.

TITLE Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :

COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primers
- Web : www.genoscope.cns.fr
BP 191 91006 EVRY cedex - FRANCE (E-mail : secreto@genoscope.cns.fr)

end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalised. Library was constructed by Life Technologies, a division of Invitrogen.

FEATURES	Location/Qualifiers
source	1. .1766

```
/organism="Homo sapiens"  
/mol_type="mRNA"
```

```
/db_xref="taxon:9606"  
/clone="CS0DG007Y118"
```

```

/tissue_type="B cells (Ramos cell line)"
/plasmid="pCMVSPORT_6"

```

ORIGIN

Query Match	97.4%;	Score 1764.4;	DB 3;	Length 1766;
Best Local Similarity	99.9%;	Pred. No. 0;		

```
Matches 1765; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

15 TGGCGGCTAGTACACCGACCTGATGAGTGGACCCAGGACCCCTCCAGTTTA 74

QY	75	GGAGCCTCCTGGGAGCTCAGAGACGTGAGCGCCCGCCCTTGAGGAGACTCCCCCATCCGG	134
Db	61	GGGACCTCCTGGGCTTCAGAGACGTGAGCGCCCGCCCTTGAGGAGACTCCCCCATCCGG	120
QY	135	GGCGGAATGATCCGGAGTCGCTCCGACAGTGTCTGTCTCTCCCTGGTCTGGGTGC	194
Db	121	GGCGGAATGATCCGGAGTCGCTCCGACAGTGTCTGTCTCTCCCTGGTCTGGGTGC	180
QY	195	AAAGTGTGGGTTCTGGGTTCTGGAATTCGGGGGCGTTCAACGATGCTGTGGCGGCT	254
Db	181	AAAGTGTGGGTTCTGGGTTCTGGAATTCGGGGGCGTTCAACGATGCTGTGGCGGCT	240
QY	255	CCTGGGATGATCCGTCGCGCGCGCGATGCGCCGGAGACGGCTTAAGCTGCGGGAGTCCGG	314
Db	241	CCTGGGATGATCCGTCGCGCGCGCGATGCGCCGGAGACGGCTTAAGCTGCGGGAGTCCGG	300
QY	315	GGCCCAAGCATTTCCGGGCTGCAATTTGACGGGGATCCCGATTCACCGCGCGCCCGC	374
Db	301	GGCCCAAGCATTTCCGGGCTGCAATTTGACGGGGATCCCGATTCACCGCGCGCCCGC	360
QY	375	GGCCCTCACCGACGGGTCCAGACCTTGATGGAGAAAGATTCGGGGACGGGTCCCTGAGAT	434
Db	361	GGCCCTCACCGACGGGTCCAGACCTTGATGGAGAAAGATTCGGGGACGGGTCCCTGAGAT	420
QY	435	CCCGATGCTTACGAGCCCAAGATGCTCAGCTTTATAGGTGTGACTTACATGTGACTTCA	494
Db	421	CCCGATGCTTACGAGCCCAAGATGCTCAGCTTTATAGGTGTGACTTACATGTGACTTCA	480
QY	495	CCTGAGTTTGTGATCCGTAATAATGGACAAATTGCAAGCTACTTCAAGTGTGTTGAGA	554
Db	481	CCTGAGTTTGTGATCCGTAATAATGGACAAATTGCAAGCTACTTCAAGTGTGTTGAGA	540
QY	555	GGATTTAAATGAAACAATGCTGTTTAAAGCTCTTTTGAGAGAGGAGCCTCGAAGCAGGGCC	614
Db	541	GGATTTAAATGAAACAATGCTGTTTAAAGCTCTTTTGAGAGAGGAGCCTCGAAGCAGGGCC	600
QY	615	TGGCCGGGAGAGACACCTGCTGTCAACGAGGAGACACAGGACGATGAAAGACCCCGTGG	674
Db	601	TGGCCGGGAGAGACACCTGCTGTCAACGAGGAGACACAGGACGATGAAAGACCCCGTGG	660
QY	675	AGCTGGCCGTGAGTGGGATGACAGACCTTCGGCCTTCAGACCGCTGCGGAGGATGCTTACC	734
Db	661	AGCTGGCCGTGAGTGGGATGACAGACCTTCGGCCTTCAGACCGCTGCGGAGGATGCTTACC	720
QY	735	GGGTCAAGGCGCAGAGCTCATATGTGTGATGAGACTCTGTTTGGACAGCCGACGAGCACCC	794
Db	721	GGGTCAAGGCGCAGAGCTCATATGTGTGATGAGACTCTGTTTGGACAGCCGACGAGCACCC	780
QY	795	GGGCTAACCCCAACCGGACTTTCGATTCGGCCCTGGGTGAGAGAGGCTTAAACGAACAAGGG	854
Db	781	GGGCTAACCCCAACCGGACTTTCGATTCGGCCCTGGGTGAGAGAGGCTTAAACGAACAAGGG	840
QY	855	TGGGCAAGAGGACATCGAAGGCTTTGGGGGCAAAAGGGAGCTGTGAGACCACTCCCTCAA	914
Db	841	TGGGCAAGAGGACATCGAAGGCTTTGGGGGCAAAAGGGAGCTGTGAGACCACTCCCTCAA	900
QY	915	GGGGCAGACACCCCACTTCACACCAAGAGAGAAACAATATCAGACCATCAGCCACA	974
Db	901	GGGGCAGACACCCCACTTCACACCAAGAGAGAAACAATATCAGACCATCAGCCACA	960
QY	975	CCCGCTTTACTGTATGATGATGCTGCTGTTGGCTCCCGATCTTAAAGGCGCACTTCGGGG	1033
Db	961	CCCGCTTTACTGTATGATGATGCTGCTGTTGGCTCCCGATCTTAAAGGCGCACTTCGGGG	1022
QY	1035	CCCGCGGATGCGCAAGGGGAGATGCGCAAAAGCTTCGTGTCTCTTGTGTGACGCCACAC	1094
Db	1021	CCCGCGGATGCGCAAGGGGAGATGCGCAAAAGCTTCGTGTCTCTTGTGTGACGCCACAC	1080
QY	1095	CTACCCCCAGGGATGACACTCGCCCGCCCAAGGAGGACACCACTTGGAGAGCATTTACAC	1155
Db	1081	CTACCCCCAGGGATGACACTCGCCCGCCCAAGGAGGACACCACTTGGAGAGCATTTACAC	1144

Accession	Source	Organism	Reference Authors Title Journal	Remark
QY 1155	CAGCTGGGCTCCCTCCAGAGACAGAGCCGGGGGACAGGGGAGACTCCACAAAGTATCTATGG			
Db 1141	CAGCTGGGCTCCCTCCAGAGACAGAGCCGGGGGACAGGGGAGACTCCACAAAGTATCTATGG			
QY 1215	GTGGGTTACACTCTTCACGCCCTCTGAAAGCGGGGACTTTCCCATTCCTCACCCACCTGA			
Db 1201	GTGGGTTACACTCTTCACGCCCTCTGAAAGCGGGGACTTTCCCATTCCTCACCCACCTGA			
QY 1275	ATGTCGCCAGACCTGGTATCATCAGCCACCAATGTCGCCCTCCACAAATGGGCTTAGAGATC			
Db 1261	ATGTCGCCAGACCTGGTATCATCAGCCACCAATGTCGCCCTCCACAAATGGGCTTAGAGATC			
QY 1335	TCAGGCTCTTCACGTCAGGGGTGACTTTCCGGAGACCCCTGGTACTTTCCAGGCTGGCT			
Db 1321	TCAGGCTCTTCACGTCAGGGGTGACTTTCCGGAGACCCCTGGTACTTTCCAGGCTGGCT			
QY 1395	CAGTTAGCATTTCACTGTCATCTACCCACGACGAGGTGGGGCCACCCAGAAACAAAGC			
Db 1381	CAGTTAGCATTTCACTGTCATCTACCCACGACGAGGTGGGGCCACCCAGAAACAAAGC			
QY 1455	CCCTTGGAAATGATTAATCTTTTATCAGGGTTGCTTATGGGGCCACGGCCGACAGTATGG			
Db 1441	CCCTTGGAAATGATTAATCTTTTATCAGGGTTGCTTATGGGGCCACGGCCGACAGTATGG			
QY 1515	CCCTTGGAGGGAGAGGAGCATTCATCAACCCAGGGAAACCCAGATTTAAAGAACCC			
Db 1501	CCCTTGGAGGGAGAGGAGCATTCATCAACCCAGGGAAACCCAGATTTAAAGAACCC			
QY 1575	CTGTGGGGGACAGACGACATAGCAGGGGTGGGCGAGTCCCTCTTATTCCTGACAAATCTC			
Db 1561	CTGTGGGGGACAGACGACATAGCAGGGGTGGGCGAGTCCCTCTTATTCCTGACAAATCTC			
QY 1635	TAGTCGATTCCTGGCTTTTCTCCCATTTGGCGATTTGGGGGCCACCTCTAAGATGCTTC			
Db 1621	TAGTCGATTCCTGGCTTTTCTCCCATTTGGCGATTTGGGGGCCACCTCTAAGATGCTTC			
QY 1695	TCTCCAGCCCTGTCTTCAACATTAATGATGTCACACCCAGGGGGCTTGGGACCTCC			
Db 1681	TCTCCAGCCCTGTCTTCAACATTAATGATGTCACACCCAGGGGGCTTGGGACCTCC			
QY 1755	CACATCATTCATTTGCTTGGCTGCCAA 1780			
Db 1741	CACATCATTCATTTGCTTGGCTGCCAA 1766			
RESULT 2				
LOCUS	CR608011			
DEFINITION	full-length cDNA clone CSDDK010YU01 of Hela cells Cot 25-normalized			
ACCESSION	CR608011			
VERSION	CR608011.1 GI:50488818			
KEYWORDS	HTC; CNSLT cDNA.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	Li, W.B., Gruber, C., Jesse, J. and Polayres, D.			
TITLE	Full-length cDNA libraries and normalization			
JOURNAL	Unpublished			
REMARK	Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/Invitrogen Corporation 1600			
REFERENCE	2 (bases 1 to 1936)			
AUTHORS	Genoscope.			
TITLE	Direct Submission			
JOURNAL	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr			
COMMENT	- Web : www.genoscope.cns.fr			
	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime			
	end enriched, double-strand cDNA was digested with Not I and cloned			
	into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library			

was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

FEATURES

source

location/Qualifiers

1..1936
/organism="Homo sapiens"
/mol_type="rRNA"
/db_xref="taxon:9606"
/clone="CSODK010Y101"
/issue_type="Hela cells Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN

Query Match 97.4%; Score 1764.4; DB 3; Length 1936;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1765; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 46 GTGGACACAGAGAGCCCTCTCCATGTTTAAAGGACCTCTGGGCTTCAGAGCGTGGCGCC 105
Db 1 GTGGACACAGAGAGCCCTCTCCATGTTTAAAGGACCTCTGGGCTTCAGAGCGTGGCGCC 60
QY 106 GGGCCCTGGGGGGAATCCCCCATCCGGGGGGGGAATGTCGGGGTGGCGGCGGCGGAGT 165
Db 61 GGGCCCTGGGGGGAATCCCCCATCCGGGGGGGGAATGTCGGGGTGGCGGCGGCGGAGT 120
QY 166 CTGCTGGCTGCTCCCTGGTGGTGGAGTGAAGTCTGGGTTCTGGAGTTTCG 225
Db 121 CTGCTGGCTGCTCCCTGGTGGTGGAGTGAAGTCTGGGTTCTGGAGTTTCG 180
QY 226 GGGCGGTTCAACGTAAGCTGTGGCGGCTCTCGGGTGAATCCGTCGGCGGGTGGCC 285
Db 181 GGGCGGTTCAACGTAAGCTGTGGCGGCTCTCGGGTGAATCCGTCGGCGGGTGGCC 240
QY 286 GGGGACGCGCTAAGGTCGCGGGGGTCCGGGGGGGCAAGATCCGGGCTTCAGATTTGAG 345
Db 241 GGGGACGCGCTAAGGTCGCGGGGGTCCGGGGGGGCAAGATCCGGGCTTCAGATTTGAG 300
QY 346 GGGATCCCGGATGACCGCGCGCCCGGCGCTTCAACGAGCGGTTCAGACTTGGTGGGA 405
Db 301 GGGATCCCGGATGACCGCGCGCCCGGCGCTTCAACGAGCGGTTCAGACTTGGTGGGA 360
QY 406 AGAAGTCCGGGAGCGGGTCCCTGAGAGATCCCGATGCTTACGAGCCAGAGTCTCAGCTT 465
Db 361 AGAAGTCCGGGAGCGGGTCCCTGAGAGATCCCGATGCTTACGAGCCAGAGTCTCAGCTT 420
QY 466 TATAGGTGACCTTCAATGATGATTTCACTCACTTTTGTGATCCGTTAAATGAGCAAA 525
Db 421 TATAGGTGACCTTCAATGATGATTTCACTCACTTTTGTGATCCGTTAAATGAGCAAA 480
QY 526 TTCGAGCTACTTCAATGATGATTTGAGAGATTAATGAACATGCTTGAAGCTCT 585
Db 481 TTCGAGCTACTTCAATGATGATTTGAGAGATTAATGAACATGCTTGAAGCTCT 540
QY 586 TTGCAAGAGGAGACCTCGAAGACAGGCGCTGGCGGCGAGAGCAACTTGTGTCAACA 645
Db 541 TTGCAAGAGGAGACCTCGAAGACAGGCGCTGGCGGCGAGAGCAACTTGTGTCAACA 600
QY 646 GACCAAGGAGCAATGAGAGACCCCGGTGAGTGGCCCTCAATGGGATGACAGACCTTGG 705
Db 601 GACCAAGGAGCAATGAGAGACCCCGGTGAGTGGCCCTCAATGGGATGACAGACCTTGG 660
QY 706 CTTTCAGACCGCTGCGAGGTGCTACCGGGTCAAGGCGAGAGCGTCAATATGTGATGA 765
Db 661 CTTTCAGACCGCTGCGAGGTGCTACCGGGTCAAGGCGAGAGCGTCAATATGTGATGA 720
QY 766 GACTCTGTTTGGCAGCGCAGAGGACCCGCGCTAACCCGACGGACTTGCATCGCGCTG 825
Db 721 GACTCTGTTTGGCAGCGCAGAGGACCCGCGCTAACCCGACGGACTTGCATCGCGCTG 780
QY 826 GGTGAGAGAGGCTAACAGAACAGAGGCTGGGAGAGAGAGATGAGAGGCTTGGGGGC 885
Db 781 GGTGAGAGAGGCTAACAGAACAGAGGCTGGGAGAGAGAGATGAGAGGCTTGGGGGC 840
QY 886 AAAAGGAGAGTGTGAGACCAACCCCTCAAGGGGCGAGACCCGACCTTACACCAAGAA 945

```

```

Db 841 AAAAGGAGAGTGTGAGACCAACCCCTCAAGGGGCGAGACCCGACCTTACACCAAGAA 900
QY 946 GAAAGAACAAATACAGACCCATAGACCAACCCCGCTTTACTGTGATGATCGCTGTTGG 1005
Db 901 GAAAGAACAAATACAGACCCATAGACCAACCCCGCTTTACTGTGATGATCGCTGTTGG 960
QY 1006 CTCGCCATCTGAAGGCGCAGGTTCGGGGGCGCGGGGATGGGAGAGGGGGAATGCCGCAA 1065
Db 961 CTCGCCATCTGAAGGCGCAGGTTCGGGGGCGCGGGGATGGGAGAGGGGGAATGCCGCAA 1020
QY 1066 GCTCCGCTGCTCTTGTGAGACGCAACCACTTACCCCGAGGGTACCACTCGCCCGCC 1125
Db 1021 GCTCCGCTGCTCTTGTGAGACGCAACCACTTACCCCGAGGGTACCACTCGCCCGCC 1080
QY 1126 CAGGAGGACCACTGCGAGCCATTCACCTGAGTCCCTCAAGACAGAGCCGGAGCC 1185
Db 1081 CAGGAGGACCACTGCGAGCCATTCACCTGAGTCCCTCAAGACAGAGCCGGAGCC 1140
QY 1186 AGCGGACAGCTCCGAGAGTTATCTATGGTGGTTACATCTTCAACGCCCTGAAAGG 1245
Db 1141 AGCGGACAGCTCCGAGAGTTATCTATGGTGGTTACATCTTCAACGCCCTGAAAGG 1200
QY 1246 GGAATTTTCCATTTCCCTTCAACCACTGATATGTCCTCAGACTGATTCAGACCAAG 1305
Db 1201 GGAATTTTCCATTTCCCTTCAACCACTGATATGTCCTCAGACTGATTCAGACCAAG 1260
QY 1306 TCCCCCAACAAATGGGCTCAGAGATCTCAGGCTTTCACGTCAGAGGAGTACCTTCCG 1365
Db 1261 TCCCCCAACAAATGGGCTCAGAGATCTCAGGCTTTCACGTCAGAGGAGTACCTTCCG 1320
QY 1366 GAGCCCTCGGATCTTCAAGGCTCGCTCAGTTGATGATTCAGAGGATTCACCCAG 1425
Db 1321 GAGCCCTCGGATCTTCAAGGCTCGCTCAGTTGATGATTCAGAGGATTCACCCAG 1380
QY 1426 ACGAGTGGGCGCAACCCAGAAACCAAGCCCTTGAATATGATCTTTTCATCAGGGT 1485
Db 1381 ACGAGTGGGCGCAACCCAGAAACCAAGCCCTTGAATATGATCTTTTCATCAGGGT 1440
QY 1441 TGCCATATGGGCGCAACCCAGAGATGAGGCTTTCAGAGGATGAGAGATCTTATCAG 1500
Db 1501 TGCCATATGGGCGCAACCCAGAGATGAGGCTTTCAGAGGATGAGAGATCTTATCAG 1460
QY 1546 CCAAGGAAACCCAGGATTAATAAGAGCCCTGTGGGGGCAAGACATAGAGGGTGG 1605
Db 1501 CCAAGGAAACCCAGGATTAATAAGAGCCCTGTGGGGGCAAGACATAGAGGGTGG 1560
QY 1606 GCAGTGCTCCCTTATCTGACAAATCTCTAGTGAATCTTGCCTTTTCTCCGATTGC 1665
Db 1561 GCAGTGCTCCCTTATCTGACAAATCTCTAGTGAATCTTGCCTTTTCTCCGATTGC 1620
QY 1666 GGATTTGGGGGCGACCTTAAAGATGCTCTCTCCAGCCCTGTCTCAACCATATCTCAAT 1725
Db 1621 GGATTTGGGGGCGACCTTAAAGATGCTCTCTCCAGCCCTGTCTCAACCATATCTCAAT 1680
QY 1726 TAGTGCACACCCAGGGGCTTGGCACTTCCCACTATGATTTTCTGCAAGTGG 1785
Db 1681 TAGTGCACACCCAGGGGCTTGGCACTTCCCACTATGATTTTCTGCAAGTGG 1740
QY 1786 AATAAAGCGGTGATTTGCCAAGCTGG 1811
Db 1741 AATAAAGCGGTGATTTGCCAAGCTGG 1766

```

RESULT 3
CR622111
LOCUS CR622111
DEFINITION full-length cDNA clone CSOD4004Y23 of Neuroblastoma of Homo
sapiens (human).
ACCESSION CR622111
VERSION CR622111.1 GI:50502918
KEYWORDS HTC; NSLT cDNA.
SOURCE Homo sapiens (human)

ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1839)
REFERENCE	Li, W.B., Gruber, C., Jessee, J. and Polayes, D. Full-length cDNA libraries and normalization Unpublished
AUTHORS	Contact : Peng Liang Email : fliang@lifetech.com URL :
JOURNAL	http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Paraday Avenue 2 (bases 1 to 1839) Genoscope.
COMMENT	Submitted (20-JUN-2004) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : sequefgenoscope.cns.fr - web : www.genoscope.cns.fr) 1st strand cDNA was primed with a NotI-oliigo(dT) primer. Five pinhead end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library has been constructed by Life Technologies, a division of Invitrogen.
FEATURES	Location/Qualifiers 1..1839 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CS0DA004YD23" /issue_type="Neuroblastoma" /plasmid="pCMVSPORT_6"
ORIGIN	
Query Match	91.6%; Score 1659; DB 3; Length 1839;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1659; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Db	153 CGCGTCGCAATGCTGCTGCTGCCCTCCCTGGTTGTGGTGCAAAAGTCTGGATTCTGGG 212
Qy	2 CGCGTCGCAATGCTGCTGCTGCCCTCCCTGGTTGTGGTGCAAAAGTCTGGATTCTGGG 61
Db	213 TTTCGTGAATTCGCGGGCCGTTACACAGTAGCCTGTGCAGCTCCTCGGGTAGTCGTC 272
Qy	62 TTTCGTGAATTCGCGGGCCGTTACACAGTAGCCTGTGCAGCTCCTCGGGTAGTCGTC 121
Db	273 GCGGCGCGGTGCCCGGGACGCGCTTAGTGTGCCGGGGGTTCGGGGCCCAAGCATTCGGG 332
Qy	122 GCGGCGCGGTGCCCGGGACGCGCTTAGTGTGCCGGGGGTTCGGGGCCCAAGCATTCGGG 181
Db	333 CTGCAGATTGACGGGGATCCCGAATGACACGCGCGCCCCCGCGCCCTCACCGAGCGGTC 392
Qy	182 CTGCAGATTGACGGGGATCCCGAATGACACGCGCGCCCCCGCGCCCTCACCGAGCGGTC 241
Db	393 AGACTGTGTGGGAAGAAGTGTGGGAGCGGGTCCCTTAGAGATCCCGATGCTTAGAGCCA 452
Qy	242 AGACTGTGTGGGAAGAAGTGTGGGAGCGGGTCCCTTAGAGATCCCGATGCTTAGAGCCA 301
Db	453 AGATGCTCAGCTTTATAGTGTGACCTACATGTAAGCTTCACTCAGTTTTGTGATCCG 512
Qy	302 AGATGCTCAGCTTTATAGTGTGACCTACATGTAAGCTTCACTCAGTTTTGTGATCCG 361
Db	513 TAAATGACAATAATTCGAGCTACTTCACAGTGTGTTGAGAGATTAAATGAACAATG 572
Qy	362 TAAATGACAATAATTCGAGCTACTTCACAGTGTGTTGAGAGATTAAATGAACAATG 421
Db	573 CTTCGTAAAGCTTTTGACAGAGAGAGCTCGGAGACAGAGGCTTCGGCGGACAGACACC 632
Qy	422 CTTCGTAAAGCTTTTGACAGAGAGAGCTCGGAGACAGAGGCTTCGGCGGACAGACACC 481
Db	633 TGCTGTACCAAGGACCAAGGACAGATGAAGACCCCGGTGAGACTGACGTCAGTGGGA 692
Qy	482 TGCTGTACCAAGGACCAAGGACAGATGAAGACCCCGGTGAGACTGACGTCAGTGGGA 541
Db	693 TGCAAGCCCTTGCGCTTCAGCACCGCTGCGAGGTGCTTACGGGTCAAGGCCAGAGCT 752

Db	542	TGCAAGACCCCTGCGCTTCAAGACCGCTGCGCGAGTGCTACCGGGTCAAGGCGAAGCGT	601
OY	753	CATATGTGATGAGACTCTGTGTTGGCAGCCGACGAGCGACCCGGCTTACCCCAACGGACT	812
Db	602	CATATGTGATGAGACTCTGTGTTGGCAGCCGAGGAGGACCCGGGCTTACCCCAACGGACT	661
OY	813	TCGATTCGGCCCTTGGGGTGGAGAAAGGCTTACAGAACAGAGGGGTGGGCAAGAGGATGGA	872
Db	662	TCGATTCGGCCCTTGGGGTGGAGAAAGGCTTACAGAAACAGAGGGGTGGGCAAGAGGATGGA	721
OY	873	AGGCGTTGGGGGCAAAAGGGAGCTGTGAGACACCCCTCAAGGGGCGACACCCCAACC	932
Db	722	AGGCGTTGGGGGCAAAAGGGAGCTGTGAGACACCCCTCAAGGGGCGACACCCCAACC	781
OY	933	TCACACCAAGGAAGAAACAATACAGACCCATACGCAACACCCCGTCTTACTGTGATG	992
Db	782	TCACACCAAGGAAGAAACAATACAGACCCATACGCAACACCCCGTCTTACTGTGATG	841
OY	993	AGTGGCTGTGTCCTCCGATCTGTGAAGCGCAGCTTGGGGGCCCGGGATGCGAAGG	105
Db	842	AGTGGCTGTGTCCTCCGATCTGTGAAGCGCAGCTTGGGGGCCCGGGATGCGAAGG	901
OY	1053	GGGATGCCGCAAAAGCTCCGTCTCTTGTGAGCGCCACACTTACCCCAAGGGTACCC	1112
Db	902	GGGATGCCGCAAAAGCTCCGTCTCTTGTGAGCGCCACACTTACCCCAAGGGTACCC	961
OY	1113	ACTGGCCCCGGCCGAGGGAGGCAACACTGGAGGCAATTACCCAGCTGGTCCCTCAAGA	1172
Db	962	ACTGGCCCCGGCCGAGGGAGGCAACACTGGAGGCAATTACCCAGCTGGTCCCTCAAGA	1021
OY	1173	CAGAGCGGGGGCAAGCGGACAGTCCCAAGAGTTATCTAAGGGTGGGTTACACTTTTAC	1232
Db	1022	CAGAGCGGGGGCAAGCGGACAGTCCCAAGAGTTATCTAAGGGTGGGTTACACTTTTAC	1081
OY	1233	GCCCCCTGGAACGGGGGACTTTCCATCCCTCAACCACTGAAGTCCCGGACACTGGATC	1292
Db	1082	GCCCCCTGGAACGGGGGACTTTCCATCCCTCAACCACTGAAGTCCCGGACACTGGATC	1141
OY	1293	ATCCAGCCACCAAGTGGCCCCCACCACAACAAATGAGGCTCAGATCTGAGGCTTCCAGTCA	1352
Db	1142	ATCCAGCCACCAAGTGGCCCCCACCACAACAAATGAGGCTCAGATCTGAGGCTTCCAGTCA	1201
OY	1353	GGGTGACCTTCCGAGGCCCTCGGTGACTTCAAGGGCTCGCTCAAGTTACATTCAGTGC	1412
Db	1202	GGGTGACCTTCCGAGGCCCTCGGTGACTTCAAGGGCTCGCTCAAGTTACATTCAGTGC	1261
OY	1413	CATCTACCCCAAGAGTGGGGCCCAACCCAGAAACCAAAGCCCCCTTGAAGATGATCT	1472
Db	1262	CATCTACCCCAAGAGTGGGGCCCAACCCAGAAACCAAAGCCCCCTTGAAGATGATCT	1321
OY	1473	CTTTTCATCAGGGTTCCTATGAGGACCAAGGAGCAAGGTATGGCCCTTGGCAGGGTACGA	1532
Db	1322	CTTTTCATCAGGGTTCCTATGAGGACCAAGGAGCAAGGTATGGCCCTTGGCAGGGTACGA	1381
OY	1533	GGAATTTCATCAGCCAGGAGAAACCCAGGTATTAAGAAAGCCCTGTGGGGGCAAGACAG	1592
Db	1382	GGAATTTCATCAGCCAGGAGAAACCCAGGTATTAAGAAAGCCCTGTGGGGGCAAGACAG	1441
OY	1593	ATAGCAGGGGTGGGACGTCCCTTATCTGACAACTCTAGTCGATTCCTTGCCCTT	1652
Db	1442	ATAGCAGGGGTGGGACGTCCCTTATCTGACAACTCTAGTCGATTCCTTGCCCTT	1501
OY	1653	TTTCCTCCGATTTGGCGAATTTGGGGGGCACTCTTAAGATGCTCTTCCAGCCCTGTCTCAA	1712
Db	1502	TTTCCTCCGATTTGGCGAATTTGGGGGGCACTCTTAAGATGCTCTTCCAGCCCTGTCTCAA	1561
OY	1713	CCATACTCAAATTAAGTGCACAACCAAGGGGCTTGGACCTTCCCATCATCATCTGTCTT	1772
Db	1562	CCATACTCAAATTAAGTGCACAACCAAGGGGCTTGGACCTTCCCATCATCATCTGTCTT	1621
OY	1773	GCTGCCAAGTGCAGATTAACGGCGTGAATTGCGCAACTGG	1811
Db	1622	GCTGCCAAGTGCAGATTAACGGCGTGAATTGCGCAACTGG	1660

[illegible]

OY	573	CTTGAAGAGCTCTTTGGACAGAGGGAGCTCCGAAACAGAGGAGCTTGGCCCGACAGACAC	632
Db	422	CTTGAAGAGCTCTTTGGACAGAGGGAGCTCCGAAAGCAGAGGAGCTTGGCCCGACAGACAC	481
OY	633	TGCTGTCAACAGGAGCAACAGGACCATGAAGACCCCGGTGAGCTGGCCGTCACTGGGA	692
Db	482	TGCTGTCAACAGGAGCAACAGGACCATGAAGACCCCGGTGAGCTGGCCGTCACTGGGA	541
OY	693	TGCGAACCCCTGGCCCTTACGACACCGCTCCGAGGTGGCTACCGGGTCAAGGTCAGAGACT	752
Db	542	TGCGAACCCCTGGCCCTTACGACACCGCTCCGAGGTGGCTACCGGGTCAAGGTCAGAGACT	601
OY	753	CATAATGTGATGAGACTCTGTTTGGCAGACCCGACAGGACACCCGGGCTACCCACCGGACT	812
Db	602	CATAATGTGATGAGACTCTGTTTGGCAGACCCGACAGGACACCCGGGCTACCCACCGGACT	661
OY	813	TCGATCCGCTGTGGGTGAGAAAGGCTTAAACAGAACAGAGCCGTGGGCAAGAGGACATCGA	872
Db	662	TCGATCCGCTGTGGGTGAGAAAGGCTTAAACAGAACAGAGCCGTGGGCAAGAGGACATCGA	721
OY	873	AGGCTTTGGGGGGCAAAAGGGAGCTGTGACACACCCCTTCAAGGGGACACACCCCAACCC	932
Db	722	AGGCTTTGGGGGGCAAAAGGGAGCTGTGACACACCCCTTCAAGGGGACACACCCCAACCC	781
OY	933	TCACACCAAGGAAAGAAACAAATACAGAACCCATCAGGCAACACCCCGTCTTACTGTGATG	992
Db	782	TCACACCAAGGAAAGAAACAAATACAGAACCCATCAGGCAACACCCCGTCTTACTGTGATG	841
OY	993	AGTCGCTGTTTGGCTCCCGATCTGAAGGCGCACGCTTCCGGGGCCCGCGGATGGCGAAG	1053
Db	842	AGTCGCTGTTTGGCTCCCGATCTGAAGGCGCACGCTTCCGGGGCCCGCGGATGGCGAAG	901
OY	1053	GGGATGCGCGCAAAAGTCCGTGCTCTTGTGTGAAGCCACACCTTACCCCAAGGGTAGCC	1113
Db	902	GGGATGCGCGCAAAAGTCCGTGCTCTTGTGTGAAGCCACACCTTACCCCAAGGGTAGCC	961
OY	1113	ACTGAGCCCGCGCCCGACAGGGAGGACCACTGCGAGGCAATTCACCCAGCTGGTCCCTCCAA	1173
Db	962	ACTGAGCCCGCGCCCGACAGGGAGGACCACTGCGAGGCAATTCACCCAGCTGGTCCCTCCAA	1022
OY	1173	CAGAGCCGGGGCCAGCCGGCAGACTTCCAGAAATTATCTATGGGTGGTTACACTCTTCA	1233
Db	1022	CAGAGCCGGGGCCAGCCGGCAGACTTCCAGAAATTATCTATGGGTGGTTACACTCTTCA	1082
OY	1233	GGCCCCCTGAAGCGGGGACTTTCCTCAATTCCTCAACCACTGTAATGTCCCAAGACTGGTC	1293
Db	1082	GGCCCCCTGAAGCGGGGACTTTCCTCAATTCCTCAACCACTGTAATGTCCCAAGACTGGTC	1142
OY	1293	ATTCAGGCAACAGGTGCCCCGCCCAACAATGGAGCCTCAGATCTCAGGCGCTTCAAGTCA	1353
Db	1142	ATTCAGGCAACAGGTGCCCCGCCCAACAATGGAGCCTCAGATCTCAGGCGCTTCAAGTCA	1202
OY	1353	GGGTGAACCTTCCGAGGCCCTCTGTGTGACTTCCAGGGCTCGCTCAGTTAGACTTTCAGTGC	1413
Db	1202	GGGTGAACCTTCCGAGGCCCTCTGTGTGACTTCCAGGGCTCGCTCAGTTAGACTTTCAGTGC	1262
OY	1413	CATCTTACCCCAACGAGGTGGGGCCACCCAGAAACCAAAACCCCTTGGAAATATACT	1473
Db	1262	CATCTTACCCCAACGAGGTGGGGCCACCCAGAAACCAAAACCCCTTGGAAATATACT	1322
OY	1473	CTTTTCATCAGGGTTCGCTATGGGGCCACGGGACAGGTATGGCCCTTGCACAGGGTAGGA	1533
Db	1322	CTTTTCATCAGGGTTCGCTATGGGGCCACGGGACAGGTATGGCCCTTGCACAGGGTAGGA	1382
OY	1533	GGACATTTCATCACCCAGGGAACCCCAAGTATTTAAAGAAAGCCCTGTGGGGGACAGAC	1593
Db	1382	GGACATTTCATCACCCAGGGAACCCCAAGTATTTAAAGAAAGCCCTGTGGGGGACAGAC	1442
OY	1593	ATTAGCAGGGGTGGGCACTGCTCCCTTATTCCTGACAACTCTCTAGTGCATTTCTTGGCTTT	1653
Db	1442	ATTAGCAGGGGTGGGCACTGCTCCCTTATTCCTGACAACTCTCTAGTGCATTTCTTGGCTTT	1501

QY	1653	TTCTCCCAATGCGGATTTGGGGGCGACCTCTAAATAGCTCTCCAGCCCTGTCTCAA	1712
Db	1502	TTCTCCCAATGCGGATTTGGGGGCGACCTCTAAATAGCTCTCTCCAGCCCTGTCTCAA	1551
QY	1713	CCATATCTCOAATTAATGATGCCAACCCAGGGGCGCTTGACCTCCACATCATCATTTGCTTT	1772
Db	1562	CCATATCTCOAATTAATGATGCCAACCCAGGGGCGCTTGACCTCCACATCATCATTTGCTTT	1621
QY	1773	GCTGCGCAAGTG 1783	
Db	1622	GCTGCGCAAGTG 1632	
RESULT 5			
CR621332			
LOCUS			
DEFINITION	CR621332	1614 bp	mRNA
			linear HTC 21-JUL-2004
			full-length cDNA clone CS0DJ008YJ14 of T cells (Jurkat cell line)
ACCESSION	CR621332		
VERSION	CR621332.1	GI:50502139	
KEYWORDS	HTC; CNSLT cDNA.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1614)		
JOURNAL	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.		
REMARK	Unpublished		
	Contact : Feng Liang Email : fliang@life.techn.com URL : http://fulllength.invitrogen.com/ invitrogen corporation 1600 Paradise Avenue 2 (bases 1 to 1614)		
REFERENCE	Genoscope.		
AUTHORS	Direct Submision		
JOURNAL	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr		
COMMENT	- Web : www.genoscope.cns.fr) 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primers end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.		
FEATURES			
Source	location/Qualifiers		
	1..1614		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CS0DJ008YJ14"		
	/tissue_type="T cells (Jurkat cell line) Cot		
	10-normalized"		
	/plasmid="pCMVSPORT_6"		
ORIGIN			
Query Match	89.1%; Score 1614; DB 3; Length 1614;		
Best Local Similarity	100.0%; Pred. No. 0;		
Matches 1614; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
QY	160	GCAAGTGTCTGCTGGCTGCTCCTCGTGTGCTGGAGTGCAGAAAGTCTGGGTTCTGGGTTTCTGG	219
Db	1	GCAAGTGTCTGCTGGCTGCTCCTCGTGTGCTGGAGTGCAGAAAGTCTGGGTTCTGGGTTTCTGG	60
QY	220	ATTGCGGGGCGGTTTACACAGTATGCTGGCGGCGCTCTCGAGTATGTCCTCGCGGCGG	279
Db	61	ATTGCGGGGCGGTTTACACAGTATGCTGGCGGCGCTCTCGAGTATGTCCTCGCGGCGG	120
QY	280	GTGCCCCGGGACGCGCTTAGGCTGCGCGGGGTTCGAGGCGCCAGGCAATTCCGGGCTGCAGA	339
Db	121	GTGCCCCGGGACGCGCTTAGGCTGCGCGGGGTTCGAGGCGCCAGGCAATTCCGGGCTGCAGA	180
QY	340	TTGAGAGGGGAGTCCCGAGATGACACCGCGCGCCCGCGCGCTTACAGAGGGGTTCAGACTGT	399
Db	181	TTGAGAGGGGAGTCCCGAGATGACACCGCGCGCCCGCGCGCTTACAGAGGGGTTCAGACTGT	240

QY	400	TTGGGAAABAGTGTGGGGAGGGGTCCCTTGAGATCCCGATGGCTACAGAGCCAGATGCT	459
Db	241	GTGGGAAABAGTGTGGGGAGGGGTCCCTTGAGATCCCGATGGCTACAGAGCCAGATGCT	300
QY	460	CAGCTTTATAGTGTGACTTACACATGTGACTTACCTCAGTTTTGTGATCCGTAAATG	519
Db	301	CAGCTTTATAGTGTGACTTACACATGTGACTTACCTCAGTTTTGTGATCCGTAAATG	360
QY	520	GACAAATTCGAAGCTACTTCAACGTGCTGTGTGAAGAGTTAAATGAACAATGTCTTGTA	579
Db	361	GACAAATTCGAAGCTACTTCAACGTGCTGTGTGAAGAGTTAAATGAACAATGTCTTGTA	420
QY	580	AGCTTTTTCAGAGAGGAGCTCGGAAGCAGAGGCTGTGCGAGAGCACACTTGCTTC	639
Db	421	AGCTTTTTCAGAGAGGAGCTCGGAAGCAGAGGCTGTGCGAGAGCACACTTGCTTC	480
QY	640	ACCAAGGACCAAGGACGATGAAAGATCCCTGTGAGCTGGCCGTCAGTGGATGACAG	699
Db	481	ACCAAGGACCAAGGACGATGAAAGATCCCTGTGAGCTGGCCGTCAGTGGATGACAG	540
QY	700	CCTGGGCTTTAGGACCGGTGCGAGGTGGCTACCGGGTCAAGGACAGAGGTCAATGT	759
Db	541	CCTGGGCTTTAGGACCGGTGCGAGGTGGCTACCGGGTCAAGGACAGAGGTCAATGT	600
QY	760	GGATGAGACTGTGTTTGGCAGCCGACGAGCAGCCCGGCTTACCCACCGGACTTGATCC	819
Db	601	GGATGAGACTGTGTTTGGCAGCCGACGAGCAGCCCGGCTTACCCACCGGACTTGATCC	660
QY	820	GGCCTGGGTGGAGAAAGGCTTACAGAACAGAGGGGTGGCAAGAGGATGGAAGGCTT	879
Db	661	GGCCTGGGTGGAGAAAGGCTTACAGAACAGAGGGGTGGCAAGAGGATGGAAGGCTT	720
QY	880	GGGGGCAAAAGGGAGCTGTGAGACCAACCCTTCAAGGGGAGCAACCCCACTTCAACC	939
Db	721	GGGGGCAAAAGGGAGCTGTGAGACCAACCCTTCAAGGGGAGCAACCCCACTTCAACC	780
QY	940	AAGGAGAAGAACAAAATACAGACCCATACGCAACCCCGTCTTACTGTGATGATGCT	999
Db	781	AAGGAGAAGAACAAAATACAGACCCATACGCAACCCCGTCTTACTGTGATGATGCT	840
QY	1000	GTTTGGCTCCCGAATCTGAAGGCGCAGCTTGAGGGGCCCGGAGATGGGGAAGGGAGTGC	1059
Db	841	GTTTGGCTCCCGAATCTGAAGGCGCAGCTTGAGGGGCCCGGAGATGGGGAAGGGAGTGC	900
QY	1060	CGAAAGCTCGTGTCTCTTGTGGAAGCAACCACTAACCCCAAGGGGTAGCCACTCGCC	1119
Db	901	CGAAAGCTCGTGTCTCTTGTGGAAGCAACCACTAACCCCAAGGGGTAGCCACTCGCC	960
QY	1120	CCGGCCCAAGGAGGACCACTGCGAGGCAATTCACCGAGCTGTGCTCCCTCAAGACAGAGCC	1179
Db	961	CCGGCCCAAGGAGGACCACTGCGAGGCAATTCACCGAGCTGTGCTCCCTCAAGACAGAGCC	1020
QY	1180	GGGGCCAGCGCAACTCCAGAAATTATCATGGGTGGGTTTACACTTTACAGCCGCCCT	1239
Db	1021	GGGGCCAGCGCAACTCCAGAAATTATCATGGGTGGGTTTACACTTTACAGCCGCCCT	1080
QY	1240	GAAGCGGGGAACTTTCCATTCCTCTCACCCCACTGAATGTCCCGAGCACTGTATCCAGC	1299
Db	1081	GAAGCGGGGAACTTTCCATTCCTCTCACCCCACTGAATGTCCCGAGCACTGTATCCAGC	1140
QY	1300	CACCAAGTACCCCAACAATAATGGGCTCAGAGATCTCAGGGCTTTCACAGTCAAGGGGTGAC	1359
Db	1141	CACCAAGTACCCCAACAATAATGGGCTCAGAGATCTCAGGGCTTTCACAGTCAAGGGGTGAC	1200
QY	1360	CTTCCGGAAGCCCTGTGTGACTTTCAGAGGCTCGCTCAGTTAGACATTTCAGTGCATCTAC	1419
Db	1201	CTTCCGGAAGCCCTGTGTGACTTTCAGAGGCTCGCTCAGTTAGACATTTCAGTGCATCTAC	1260
QY	1420	CCCAACAGAGGTGGGCGACCCAGAAACCAAAACCCCTTGGAATGAATCTCTTTTCAT	1479
Db	1261	CCCAACAGAGGTGGGCGACCCAGAAACCAAAACCCCTTGGAATGAATCTCTTTTCAT	1320

QY	1480	CAGGGTTGCTATGGGGGACAGGGGACAGGATATGGCCCTTGCCAGGGGTAGAGACACT	1539		
Db	1321	CAGGGTCTCTATGGGGGACAGGGGACAGGATATGGCCCTTGCCAGGGGTAGAGACACT	1380		
QY	1540	CATCACCCAGGGAACCCGAGTATTAAGAACCCCTGTGGGGGACAGACACATATGACG	1599		
Db	1381	CATCACCCAGGGAACCCGAGTATTAAGAACCCCTGTGGGGGACAGACACATATGACG	1440		
QY	1600	GGGTGGGACATGTCCTCTCTTATATCTGACAAATCTCTAGTCATCTTGGCTTTTCTCC	1659		
Db	1441	GGGTGGGACATGTCCTCTCTTATATCTGACAAATCTCTAGTCATCTTGGCTTTTCTCC	1500		
QY	1660	GATGCGGATTTGGGGGGGACACCTCAATATGCTCTCTCCAGCCCTGTCTCAACATACT	1719		
Db	1501	GATGCGGATTTGGGGGGGACACCTCTTAATATGCTCTCTCCAGCCCTGTCTCAACATACT	1560		
QY	1720	CCAAATTATGTCGAACCCAGGGGCGCTGGCACCTCCACATCATCATTTGTCTTG	1773		
Db	1561	CCAAATTATGTCGAACCCAGGGGCGCTGGCACCTCCACATCATCATTTGTCTTG	1614		
RESULT 6					
CR600183		1780 bp	mRNA	linear	HTC 21-JUL-2004
LOCUS					
DEFINITION		full-length cDNA clone CS0DC011YN03 of Neuroblastoma Cot			
ACCESSION		25-normalized of Homo sapiens (human).			
VERSION		CR600183			
KEYWORDS		CR600183.1 GI:50480990			
SOURCE		HTC; CNSLT_cDNA.			
ORGANISM		Homo sapiens (human)			
		Homo sapiens			
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE		1 (bases 1 to 1780)			
AUTHORS		Li, W.B., Gruber, C., Jessee, J. and Polayes, D.			
TITLE		Full-length cDNA libraries and normalized			
JOURNAL		Unpublished			
REMARK		Contact : Peng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue 2 (bases 1 to 1780)			
REFERENCE		Genoscope.			
AUTHORS		Direct Submisison			
TITLE		Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :			
JOURNAL		BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr			
COMMENT		- Web : www.genoscope.cns.fr)			
		1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime			
		end enriched, double-strand cDNA was digested with Not I and cloned			
		into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library			
		was normalized. Library was constructed by Life Technologies, a			
		division of Invitrogen.			
FEATURES					
source		Location/Qualifiers			
		1..1780			
		/organism="Homo sapiens"			
		/mol_type="mRNA"			
		/db_xref="taxon:9606"			
		/clone="CS0DC011YN03"			
		/tissue_type="Neuroblastoma Cot 25-normalized"			
		/plasmid="pCMVSPORT_6"			
ORIGIN					
Query Match		88.3% ; Score 1600 ; DB 3 ; Length 1780 ;			
Best Local Similarity		100.0% ; Pred. No. 0 ;			
Matches 1600 ;		Conservative 0 ; Mismatches 0 ; Indels 0 ; Gaps 0 ;			
QY	212	GTTCCTGATTCGCGGGCGCTTCAACAGTACCTGTGCGGCTCTCGGGTGAATCCGTC	271		
Db	1	GTTCCTGATTCGCGGGCGCTTCAACAGTACCTGTGCGGCTCTCGGGTGAATCCGTC	60		
QY	272	CGCGGCGCGTCCCCCGGAGACGGCTTAGGCTCCGGGGGTCCGGGGGCCCAAGCAATTCGG	331		
Db	61	CGCGGCGCGTCCCCCGGAGACGGCTTAGGCTCCGGGGGTCCGGGGGCCCAAGCAATTCGG	120		

Qy	332	GCTCAGATTGACGGGGATCCCGATATGACCGCGCGCCCGCGGCTTCAACGAGGGTC	391
Dp	121	GCTCAGATTGACGGGGATCCCGATATGACCGCGCGCCCGCGCTTCAACGAGGGTC	180
Qy	392	CAGACCTGTGGGAAGAGTGCGGGGACGGGTCCCTAGAGATCCCAATGCTTACGAGCC	451
Dp	181	CAGACCTGTGGGAAGAGTGCGGGGACGGGTCCCTAGAGATCCCAATGCTTACGAGCC	240
Qy	452	AAGATGCTCAGCTTATATAGTGTGACCTTACATGTGACTTCACTCAGTTTGTGATCC	511
Dp	241	AAGATGCTCAGCTTATATAGTGTGACCTTACATGTGACTTCACTCAGTTTGTGATCC	300
Qy	512	GTAAATATGACAAATTTGCAACTCTTACAGTGTCTTTGAGAGATTAAATGAAACAAT	571
Dp	301	GTAAATATGACAAATTTGCAACTCTTACAGTGTCTTTGAGAGATTAAATGAAACAAT	360
Qy	572	GCTTGTAAAGCTCTTTGACGAGGAGGAGCTGTGGAAAGCAGGGGCTGTGGCGGACACAC	631
Dp	361	GCTTGTAAAGCTCTTTGACGAGGAGGAGCTGTGGAAAGCAGGGGCTGTGGCGGACACAC	420
Qy	632	CTGTGTCAACGAGGAGCACAGGACATGAAACCCTGTGAGACTGTGACGTGAGTGGG	691
Dp	421	CTGTGTCAACGAGGAGCACAGGACATGAAACCCTGTGAGACTGTGAGCTGAGTGGG	480
Qy	692	ATGACAGACCTTGGGCTTACAGACCGCGTCCGAGGTGTCTACCGGGTCAAGGCCAGAGC	751
Dp	481	ATGACAGACCTTGGGCTTACAGACCGCGTCCGAGGTGTCTACCGGGTCAAGGCCAGAGC	540
Qy	752	TCAATATGTGATGATGACTCTGTGTTGGAGCCGACAGGACCCGGGCTACCCCAACGGAC	811
Dp	541	TCAATATGTGATGATGACTCTGTGTTGGAGCCGACAGGACCCGGGCTACCCCAACGGAC	600
Qy	812	TTTCGATCCGCTTGGTGTGAGAGGCTTAAAGAAACAGAGGCGTGTGGCAAGGACATCG	871
Dp	601	TTTCGATCCGCTTGGTGTGAGAGGCTTAAAGAAACAGAGGCGTGTGGCAAGGACATCG	660
Qy	872	AAGGCTTTGGGGGCAAAAGGGAGCTGTGAGACCACTCCCTTCAAGGGGACGACCCCCACC	931
Dp	661	AAGGCTTTGGGGGCAAAAGGGAGCTGTGAGACCACTCCCTTCAAGGGGACGACCCCCACC	720
Qy	932	CTCACACCAAGGAAGAAACAATATACAGACCCATCAGCCCAACCCGCTTATCTGTAT	991
Dp	721	CTCACACCAAGGAAGAAACAATATACAGACCCATCAGCCCAACCCGCTTATCTGTAT	780
Qy	992	GAGTGCCTGTGTGCTCCCGATCTGAAAGGCGCCAGCTTCCGGGGCCCCCGGATGTGCAAG	1051
Dp	781	GAGTGCCTGTGTGCTCCCGATCTGAAAGGCGCCAGCTTCCGGGGCCCCCGGATGTGCAAG	840
Qy	1052	GGGGATGCGCGAAAGCTCCGTCTCTTGTGTGACGCCACCACTTACCCCGAGGGGTAGC	1111
Dp	841	GGGGATGCGCGAAAGCTCCGTCTCTTGTGTGACGCCACCACTTACCCCGAGGGGTAGC	900
Qy	1112	CACCTGCCCCCGCCCCCAGGAGGACCACTGCGAGCATTTACCCAGCTGTGTCCTTCAAG	1177
Dp	901	CACCTGCCCCCGCCCCCAGGAGGACCACTGCGAGCATTTACCCAGCTGTGTCCTTCAAG	960
Qy	1172	ACAGAGCGGGGGCCAGGCGGACGACTCCCAAGAAATTATGTAGGGGTGGTTACACTTTCA	1231
Dp	961	ACAGAGCGGGGGCCAGGCGGACGACTCCCAAGAAATTATGTAGGGGTGGTTACACTTTCA	1022
Qy	1232	CGCCCCCTGAAAGCGGGGACTTTTCCCATTTCCCTCACCCACTGAATGTGCCAGCACTGT	1291
Dp	1021	CGCCCCCTGAAAGCGGGGACTTTTCCCATTTCCCTCACCCACTGAATGTGCCAGCACTGT	1080
Qy	1292	CATTCACGACCAAGTGCCTCCCAACAACAATGTGGGCTCAGAAATTCAGAGCTTCCACGTCA	1351
Dp	1081	CATTCACGACCAAGTGCCTCCCAACAACAATGTGGGCTCAGAAATTCAGAGCTTCCACGTCA	1140
Qy	1352	GGGGTGAACCTTCCGGAGGCCCTCTGTGTGACTTCAAGGGCTGACTCAGTTAGACATTTCAAGT	1411
Dp	1141	GGGGTGAACCTTCCGGAGGCCCTCTGTGTGACTTCAAGGGCTGACTCAGTTAGACATTTCAAGT	1200
Qy	1412	CCATCTTACCCACGACGAGGTGGGGCACCCAGAAACCAAAAGCCCTTGTGAAATGATATAC	1471

Db	1201	CCATCTACCCCAAGAGGTGGGGCCACCAGAAACCAAGCCCCCTTGGAAATGATAC	1260
OY	1472	TCTTTCAATCAGGGTTCCTATGGGGCCACGGGACAGGTATGGCCCTTGCAGGGTAGG	1531
Db	1261	TCTTTCAATCAGGGTTCCTATGGGGCCACGGGACAGGTATGGCCCTTGCAGGGTAGG	1320
OY	1532	AGGACATTCATCACCCAGGGAAACCCAGGTATTAAAGAAAGCCCTGGGGGGCAGACAGA	1531
Db	1331	AGGACATTCATCACCCAGGGAAACCCAGGTATTAAAGAAAGCCCTGTGGGGGCAGACAGA	1380
OY	1532	CATACAGAGGGGTGGCAGTGCCCTCCCTTATCTGACAAATCTATAGTGATTTCTGCCTT	1651
Db	1381	CATACAGAGGGGTGGCAGTGCCCTCCCTTATCTGACAAATCTATAGTGATTTCTGCCTT	1440
OY	1652	TTTCTCCCGATTGGCGGATTTGGGGGGCAACTTAGAGGCTCTCTCCAGGCCGTGTCCA	1711
Db	1441	TTTCTCCCGATTGGCGGATTTGGGGGGCAACTTAGAGGCTCTCTCCAGGCCGTGTCCA	1500
OY	1712	ACCATATCTCCAATTATAGTGCCAACCCAGAGGGGCTGGCACTCCGACATCATCATATTGCTT	1711
Db	1501	ACCATATCTCCAATTATAGTGCCAACCCAGAGGGGCTGGCACTCCGACATCATCATATTGCTT	1560
OY	1772	TGCTGCCAAGTGCAGATTAAAGGGGTGATTGGCAACTGG	1811
Db	1561	TGCTGCCAAGTGCAGATTAAAGGGGTGATTGGCAACTGG	1600

Db 1 GCTGCTCCTTGTTCTGGGTGCAGATGCTGGATTCTGGGTTTCTGGATTTCGAGTTCCGCGCCG 60

QY 232 TTCAACAGTAGCTGTGCGCGGCTCCTCGGATGAGTCGTCGCGCGCGGATGCCCGGAGC 291

Db 61 TTCAACAGTAGCTGTGCGCGGCTCCTCGGATGAGTCGTCGCGCGCGGATGCCCGGAGC 120

QY 292 GGCCTAAGGCTGCGGGGGGTCCGGGGGCCCAAGGCAATCCGGGCTGAGATTGACGGGGATC 351

Db 121 GGCCTAAGGCTGCGGGGGGTCCGGGGGCCCAAGGCAATCCGGGCTGAGATTGACGGGGATC 180

QY 352 CGGATGACCGCGCGCCCGCGGCTCAACGAGCGGATCCAGACTGATGGAGAAAG 411

Db 181 CGGATGACCGCGCGCCCGCGGCTCAACGAGCGGATCCAGACTGATGGAGAAAG 240

QY 412 TGCGGGGACGGGTCCTTAGAGATCCCATGCTTAGAGACCAAGATGCTCAAGTTATAG 471

Db 241 TGCGGGGACGGGTCCTTAGAGATCCCATGCTTAGAGACCAAGATGCTCAAGTTATAG 300

QY 472 TGTGACCAACATGTGACTCACTCACTGATTTTGATCCGTAATAATGACAAATTCGAA 531

Db 301 TGTGACCAACATGTGACTCACTCACTGATTTTGATCCGTAATAATGACAAATTCGAA 360

QY 532 GCTACTTCAACGTGCTGTAGAGAGATTAAATGAACAATGCTTTGAAGCTCTTTCAG 591

Db 361 GCTACTTCAACGTGCTGTAGAGAGATTAAATGAACAATGCTTTGAAGCTCTTTCAG 420

LOCUS	CR597629	1596 bp	mRNA	linear	HTC 21-JUL-2004
DEFINITION	Full-length cDNA clone CS0DL002Y105 of B cells (Ramos cell line)				
ACCESSION	CR597629	Cot 25-normalized of Homo sapiens (human).			
VERSION	CR597629.1	GI:50478436			
KEYWORDS	HTC; cNSLT cDNA.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 1596)				
TITLE	Li, W.B., Gruber, C., Jesssee, J. and Polayes, D.				
JOURNAL	Full-length cDNA libraries and normalization				
REMARK	Unpublished				
REFERENCE	Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/ Invitrogen Corporation 1600				
AUTHORS	Faraday Avenue				
TITLE	2 (bases 1 to 1596)				
JOURNAL	Genoscope.				
COMMENT	Direct Submission				
	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :				
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr				
	- Web : www.genoscope.cns.fr)				
	1st strand cDNA was primed with a NotI-oligo(dAT) primer. Five prime				
	ends enriched, double-strand cDNA was digested with Not I and cloned				
	into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library				
	was normalized. Library was constructed by Life Technologies, a				
	division of Invitrogen.				
FEATURES	Location/Qualifiers				
Source	1..1596				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="CS0DL002Y105"				
	/tissue_type="B cells (Ramos cell line) Cot 25-normalized"				
	/plasmid="pCMVSPORT_6"				
ORIGIN	88.1%; Score 1596; DB 3; Length 1596;				
Query Match	Best Local Similarity 100.0%; Pred. No. 0;				
Matches 1596;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;				

421	GAGGAGCCTCGAAGACAGAGGCGTGGCCGAGAGCAACCTGGCTGTCAACAGGAGCAAC	480
481	AGGCAGCATGAAGACCCCGGTGAGCTGGCCGCTCAGTGGGAATGCAGACCCCTCGGCTTCA	711
541	AGGCAGCATGAAGACCCCGGTGAGCTGGCCGCTCAGTGGGAATGCAGACCCCTCGGCTTCA	540
712	GCACCGCTGGCCGAGGTGGCTTACCCGGTTCAAAGCCAGAGACGTCAATATGAGATGAGACTCT	771
832	GCACCGCTGGCCGAGGTGGCTTACCCGGTTCAAAGCCAGAGACGTCAATATGAGATGAGACTCT	600
952	GCACCGCTGGCCGAGGTGGCTTACCCGGTTCAAAGCCAGAGACGTCAATATGAGATGAGACTCT	600
1012	GTTTGGCAGCCAGCAGGCAACCCCGCTTACCCCAACCGACTTTCATCCGCTCGGTGGTGA	831
1072	GTTTGGCAGCCAGCAGGCAACCCCGCTTACCCCAACCGACTTTCATCCGCTCGGTGGTGA	660
1132	GTTTGGCAGCCAGCAGGCAACCCCGCTTACCCCAACCGACTTTCATCCGCTCGGTGGTGA	660
1192	GAAAGCTTAAAGAAACCAAGAGCGCTGGGCTAAAGAGGCTTCAAGAGGCTTTGGGGGCAAAAGG	891
1252	GAAAGCTTAAAGAAACCAAGAGCGCTGGGCTAAAGAGGCTTCAAGAGGCTTTGGGGGCAAAAGG	720
1311	GAAAGCTTAAAGAAACCAAGAGCGCTGGGCTAAAGAGGCTTCAAGAGGCTTTGGGGGCAAAAGG	951
1371	GAGCTGTGAGACACACCCCTCAAGGGGAGAGACCCCAACCTTCACACCAAGAAAGAA	951
1431	GAGCTGTGAGACACACCCCTCAAGGGGAGAGACCCCAACCTTCACACCAAGAAAGAA	780
1491	GAGCTGTGAGACACACCCCTCAAGGGGAGAGACCCCAACCTTCACACCAAGAAAGAA	780
1551	CAAAATACAGACCCATCAGCCACACCCCGCTTTCATCTGTGATGAGTGCCTGTTTGGCTCCG	1011
1611	CAAAATACAGACCCATCAGCCACACCCCGCTTTCATCTGTGATGAGTGCCTGTTTGGCTCCG	840
1671	CAAAATACAGACCCATCAGCCACACCCCGCTTTCATCTGTGATGAGTGCCTGTTTGGCTCCG	840
1731	ATCTGAAGAGGCGCAGCTTCGCGGGGCGCGGAGTGGCGAAAGGGGGATGCGCGAAAGCTCG	1071
1791	ATCTGAAGAGGCGCAGCTTCGCGGGGCGCGGAGTGGCGAAAGGGGGATGCGCGAAAGCTCG	900
1851	ATCTGAAGAGGCGCAGCTTCGCGGGGCGCGGAGTGGCGAAAGGGGGATGCGCGAAAGCTCG	900
1911	TGCTCTCTTGTGAGCGCCACCACTTACCCCAAGGGGTAGCCACTCGCCCGCCCAAGGA	1131
1971	TGCTCTCTTGTGAGCGCCACCACTTACCCCAAGGGGTAGCCACTCGCCCGCCCAAGGA	960
2031	TGCTCTCTTGTGAGCGCCACCACTTACCCCAAGGGGTAGCCACTCGCCCGCCCAAGGA	960
2091	TGCTCTCTTGTGAGCGCCACCACTTACCCCAAGGGGTAGCCACTCGCCCGCCCAAGGA	1191
2151	TGCTCTCTTGTGAGCGCCACCACTTACCCCAAGGGGTAGCCACTCGCCCGCCCAAGGA	1020
2211	GAGCTCCAGAAATATCTATGAGGTGGGTTCACCTTTCACGCGCCCTGAAGCGGGAGCT	1251
2271	GAGCTCCAGAAATATCTATGAGGTGGGTTCACCTTTCACGCGCCCTGAAGCGGGAGCT	1080
2331	GAGCTCCAGAAATATCTATGAGGTGGGTTCACCTTTCACGCGCCCTGAAGCGGGAGCT	1311

Db 1081 TTCACATTCCTCCACCCACCTGAAATGTCCCCAGACCTGTCTCATCCAGCCACCAAGTGCACC 1140
QY 1312 CCACCAATATGGGCTCAGGATCTCAGGCTTCCAGTACGGGTGACCTTCCGAGCCC 1371
Db 1141 CCACCAATATGGGCTCAGGATCTCAGGCTTCCAGTACGGGTGACCTTCCGAGCCC 1200
QY 1372 CTTGTGATCTCAGGCTCAGGCTTCCAGTACGGGTGACCTTCCGAGCCC 1431
Db 1201 CTTGTGATCTCAGGCTCAGGCTTCCAGTACGGGTGACCTTCCGAGCCC 1260
QY 1432 TGGGGCCACCCAGAAACCAAGCCCCCTTGAAGATGATCTTTTCATCAGGGGTGCTTA 1491
Db 1261 TGGGGCCACCCAGAAACCAAGCCCCCTTGAAGATGATCTTTTCATCAGGGGTGCTTA 1320
QY 1492 TGGGGCCACCCAGAAACCAAGCCCCCTTGAAGATGATCTTTTCATCAGGGGTGCTTA 1551
Db 1321 TGGGGCCACCCAGAAACCAAGCCCCCTTGAAGATGATCTTTTCATCAGGGGTGCTTA 1380
QY 1552 AACCCAGGATTTAAAGAAAGCCCTGTGGGGGAGACAGACATAGAGGGGTGGGAGTG 1611
Db 1381 AACCCAGGATTTAAAGAAAGCCCTGTGGGGGAGACAGACATAGAGGGGTGGGAGTG 1440
QY 1612 CCTCCCTTATCTGACAAATCTCTAGTGCATTTCTGCTTTTCTCCGATTTGCGATTT 1671
Db 1441 CCTCCCTTATCTGACAAATCTCTAGTGCATTTCTGCTTTTCTCCGATTTGCGATTT 1500
QY 1672 GGGGGCCACCTTAAGATGCTCTCTCCAGCCCTGTCTACACATACTCAATTAATAGTGC 1731
Db 1501 GGGGGCCACCTTAAGATGCTCTCTCCAGCCCTGTCTACACATACTCAATTAATAGTGC 1560
QY 1732 CAACCCAGGGGCTGAGACCTCCCAATCATCATTT 1767
Db 1561 CAACCCAGGGGCTGAGACCTCCCAATCATCATTT 1596

RESULT 8
CR606890 1575 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CL0BB006ZC02 of Neuroblastoma of Homo
DEFINITION sapiens (human).
ACCESSION CR606890.1 GI:50487697
VERSION CR606890.1
KEYWORDS HTC; cDNA; cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1575)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact: Feng Liang Email: fliang@life.technet.com URL:
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paradise Avenue
2 (bases 1 to 1575)
REFERENCE Genoscope.
AUTHORS Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail: sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source location/Qualifiers
1..1575
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CL0BB006ZC02"
/issue_type="Neuroblastoma"
/plasmid="pCMVSPORT_6"

ORIGIN

Query Match 86.0%; Score 1557; DB 3; Length 1575;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1575; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 213 TTTTCGATTTGCGGGGCGGTTACACAGTACCTGTGCGGGCTCTGCGGGTGAAGTCCGTC 272
Db 1 TTTTCGATTTGCGGGGCGGTTACACAGTACCTGTGCGGGCTCTGCGGGTGAAGTCCGTC 60
QY 273 GCGCGCGGTGCCCGGAGCGGCTTAGGCTGCGGGGGTCCGGGCGCCAGCATTCGGG 332
Db 61 GCGCGCGGTGCCCGGAGCGGCTTAGGCTGCGGGGGTCCGGGCGCCAGCATTCGGG 120
QY 333 CTGCAATTTGAAGGGGATCCCGATGACCGCGCGCCCGGCTTCAACCGAGGTCC 392
Db 121 CTGCAATTTGAAGGGGATCCCGATGACCGCGCGCCCGGCTTCAACCGAGGTCC 180
QY 393 AGACCTGGTGGGAAGAGGTGCGGGAGCGGGTCCCTGAAGATCCGATGCTACAGGCA 452
Db 181 AGACCTGGTGGGAAGAGGTGCGGGAGCGGGTCCCTGAAGATCCGATGCTACAGGCA 240
QY 453 AGATGCTAGCTTTTATGAGTGTACCTACATGTGACTTCACTCAGTTTGTGATCCG 512
Db 241 AGATGCTC-----AGGTGTACCTACATGTGACTTCACTCAGTTTGTGATCCG 292
QY 513 TAAATGGAACAATTTGAAAGTACTTACAGTGTCTGTGAGAGATTAAATGAACATG 572
Db 293 TAAATGGAACAATTTGAAAGTACTTACAGTGTCTGTGAGAGATTAAATGAACATG 352
QY 573 CTGTAAGCTCTTTGAGAGAGGAGCTCCGGAACAGAGGCTGCGCGGAGAGCAACAC 632
Db 353 CTGTAAGCTCTTTGAGAGAGGAGCTCCGGAACAGAGGCTGCGCGGAGAGCAACAC 412
QY 633 TGGTGTCAACAGGAGCAACAGGACATGAAAGACCCCGTGAAGTGCCTCACTGAGGA 692
Db 413 TGGTGTCAACAGGAGCAACAGGACATGAAAGACCCCGTGAAGTGCCTCACTGAGGA 472
QY 693 TGCAGACCTTGGGCTTACAGCACCGCTGCGGAGTGGCTACCGGGTCAAGGCGAGAGCT 752
Db 473 TGCAGACCTTGGGCTTACAGCACCGCTGCGGAGTGGCTACCGGGTCAAGGCGAGAGCT 532
QY 753 CATATGATGATGAGCTCTGTTTGGACGCCAGAGGACCCCGGCTACCCCAACGAGT 812
Db 533 CATATGATGATGAGCTCTGTTTGGACGCCAGAGGACCCCGGCTACCCCAACGAGT 592
QY 813 TCGATCCGCTGCTGAGGAGAGGCTTAACAGAACAGAGGCTGTGGCAAGAGGATCGA 872
Db 593 TCGATCCGCTGCTGAGGAGAGGCTTAACAGAACAGAGGCTGTGGCAAGAGGATCGA 652
QY 873 AGGCTTGGGGCAAAAGGGAGCTGTGAGACCCCTCAAGGGGAGAGACCCCAACCC 932
Db 653 AGGCTTGGGGCAAAAGGGAGCTGTGAGACCCCTCAAGGGGAGAGACCCCAACCC 712
QY 933 TCACACCAAGGAAGAAAGAAATAACAGACCATAGACACACCCCGTTACTGTGATG 992
Db 713 TCACACCAAGGAAGAAAGAAATAACAGACCATAGACACACCCCGTTACTGTGATG 772
QY 993 AGTCCTGTTTGGCTCCGATCTGAAGGCGCAGCTTGGGGGCGCGGAGATGGGAAG 1052
Db 773 AGTCCTGTTTGGCTCCGATCTGAAGGCGCAGCTTGGGGGCGCGGAGATGGGAAG 832
QY 1053 GGGATGCGCAAAAGCTCGTGTCTCTTGTGAGCGCACACCTAACCCCGAGGGGTAGCC 1112
Db 833 GGGATGCGCAAAAGCTCGTGTCTCTTGTGAGCGCACACCTAACCCCGAGGGGTAGCC 892
QY 1113 ACTGCGCGCGCGCCAGAGAGACACATGCGAGCATTCACCCAGCTGTCCCTCAAGA 1172
Db 893 ACTGCGCGCGCGCCAGAGAGACACATGCGAGCATTCACCCAGCTGTCCCTCAAGA 952
QY 1173 CAGAGCCGGGCGCAGCGGACACTCCAGAGTTATCTATGAGTGGTTACACTCTTAC 1232
Db 953 CAGAGCCGGGCGCAGCGGACACTCCAGAGTTATCTATGAGTGGTTACACTCTTAC 1012

QY 1233 GCCCCTGAAGCGGGAGCTTTCCATTCCTCACCCACTGATGTCCCGACAGCTGGTC 1292
Db 1013 GCCCCTGAAGCGGGAGCTTTCCATTCCTCACCCACTGATGTCCCGACAGCTGGTC 1072
QY 1293 ATCCAGCCAGCAGTGGCCCCCACAATAATGGGCTCAGAGTCTGAGGCTTTCAGCTCAG 1352
Db 1073 ATCCAGCCAGCAGTGGCCCCCACAATAATGGGCTCAGAGTCTGAGGCTTTCAGCTCAG 1132
QY 1353 GGGGACCTTCCGAGAGCCCCCTGGTGACTTCCAGGGCTCGCTCAGTTAGCATTTCACTGC 1412
Db 1133 GGGGACCTTCCGAGAGCCCCCTGGTGACTTCCAGGGCTCGCTCAGTTAGCATTTCACTGC 1192
QY 1413 CATTCACCCCAAGCAGAGTGGGCGCAACCCAGAAACCAAGCCCCCTTGGAAATGATACT 1472
Db 1193 CATTCACCCCAAGCAGAGTGGGCGCAACCCAGAAACCAAGCCCCCTTGGAAATGATACT 1252
QY 1473 CTTTCATCAGGGTGGCTTATGGGCGCAAGGAGATGGCCCTTGGCAGGGTGAAGA 1532
Db 1253 CTTTCATCAGGGTGGCTTATGGGCGCAAGGAGATGGCCCTTGGCAGGGTGAAGA 1312
QY 1533 GGACATTCATCAGCCAGGGAACCCAGGTATTAAGAAGCCCCCTGGGCGGACAGAC 1592
Db 1313 GGACATTCATCAGCCAGGGAACCCAGGTATTAAGAAGCCCCCTGGGCGGACAGAC 1372
QY 1593 ATACAGAGGGGTGGCAGTGCCTCCCTTATCTGACATCTCTAGTGAATTTCTGGCTTT 1652
Db 1373 ATACAGAGGGGTGGCAGTGCCTCCCTTATCTGACATCTCTAGTGAATTTCTGGCTTT 1432
QY 1653 TTCTCCGATTTGGGATTTGGGCGCACTCTTAAGATGCTCTCTCAAGCCCTGTCTCA 1712
Db 1433 TTCTCCGATTTGGGATTTGGGCGCACTCTTAAGATGCTCTCTCAAGCCCTGTCTCA 1492
QY 1713 CCATCTCCAAATTAAGTGCACACCCAGGGCGTGGCACTCCCAATCATCATCTGCTT 1772
Db 1493 CCATCTCCAAATTAAGTGCACACCCAGGGCGTGGCACTCCCAATCATCATCTGCTT 1552
QY 1773 GCTGCCAAGTGCATTAACCGGC 1795
Db 1553 GCTGCCAAGTGCATTAACCGGC 1575

RESULT 9
CR625812
LOCUS 1560 bp mRNA linear HTC 21-JUL-2004
DEFINITION full-length cDNA clone CSOD1019Y108 of Placenta Cot 25-normalized
of Homo sapiens (human).
ACCESSION CR625812
VERSION CR625812.1 GI:50506619
KEYWORDS HTC; CNSLT cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1560)
AUTHORS La,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue
2 (bases 1 to 1560)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers

source 1..1560
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSOD1019Y108"
/issue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 85.1%; Score 1541; DB 3; Length 1560;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1560; Conservative 0; Mismatches 0; Indels 9; Gaps 1;
QY 191 GTGCAAGTCTGGGTTCTGAGTTTCTGATTCGCGGGCCGTTCAACGAGCTGTGCC 250
Db 1 GTGCAAGTCTGGGTTCTGAGTTTCTGATTCGCGGGCCGTTCAACGAGCTGTGCC 60
QY 251 GGGTCTCGGGTGAATCCGTCGCGCGCGGTGCCCGGACCGGCTTGAAGCTGCGGGGT 310
Db 61 GGGTCTCGGGTGAATCCGTCGCGCGCGGTGCCCGGACCGGCTTGAAGCTGCGGGGT 120
QY 311 CCGGGGCCCCAGGCAATTCGGGCTGCAAGATTGACGGGATCCCGGATGACCCGCGGCC 370
Db 121 CCGGGGCCCCAGGCAATTCGGGCTGCAAGATTGACGGGATCCCGGATGACCCGCGGCC 180
QY 371 CCGGCGCCTCACGAGCGGGTCCAGACCTGGTGGGAAGAAAGTGCGGGGAACGGTCCCTGA 430
Db 181 CCGGCGCCTCACGAGCGGGTCCAGACCTGGTGGGAAGAAAGTGCGGGGAACGGTCCCTGA 240
QY 431 GGATCCGATGCTTACGAGCGCAAGATGCTTATAGTGTGACTTACACATGTGAC 490
Db 241 GGATCCGATGCTTACGAGCGCAAGATGCTTATAGTGTGACTTACACATGTGAC 300
QY 491 TTCACTCAATTTTGTGATCCGTAATTAAGCAAAATTCGAAGTACTTCAAGTGTGTT 550
Db 301 TTCACTCAATTTTGTGATCCGTAATTAAGCAAAATTCGAAGTACTTCAAGTGTGTT 350
QY 551 GAGAGATTAAATGAACCAATGCTTTGAAAGCTTTTGGAGAGGAGGCTCGGAAGCAG 610
Db 361 GAGAGATTAAATGAACCAATGCTTTGAAAGCTTTTGGAGAGGAGGCTCGGAAGCAG 420
QY 611 GGCCTGGCGGCGAGACCACTCTGTGTCAACGAGCAACAGGACGATGAAGACCCCC 670
Db 421 GGCCTGGCGGCGAGACCACTCTGTGTCAACGAGCAACAGGACGATGAAGACCCCC 480
QY 671 GTGAGCTGGCGGTGAGTGGAGTGCAGACCTCGCTTCAAGACCGCTGCGAGGTGGC 730
Db 481 GTGAGCTGGCGGTGAGTGGAGTGCAGACCTCGCTTCAAGACCGCTGCGAGGTGGC 540
QY 731 TACCGGATCAAGGCGCAGACGTCATATGATGAAGTCTGTTTGGAGCCGACAGGCG 790
Db 541 TACCGGATCAAGGCGCAGACGTCATATGATGAAGTCTGTTTGGAGCCGACAGGCG 660
QY 791 ACCCGGCTTACCCGACCGGACTTGATCCGCTTGGGTGAGAGGCTTACGAGCCAG 850
Db 601 ACCCGGCTTACCCGACCGGACTTGATCCGCTTGGGTGAGAGGCTTACGAGCCAG 660
QY 851 GCGCTGGGCAAGGAGCATCGAAGGCTTGGGGGCAAGGGGAGCTGTAGACCAACCCC 910
Db 661 GCGCTGGGCAAGGAGCATCGAAGGCTTGGGGGCAAGGGGAGCTGTAGACCAACCCC 720
QY 911 TCAAGGGGAGCAGACCCCAACCTTCAACCAAGAGAGAGCAAAATACAGACCCATCAGC 970
Db 721 TCAAGGGGAGCAGACCCCAACCTTCAACCAAGAGAGAGCAAAATACAGACCCATCAGC 771
QY 971 CACACCCCGTCTTACTGTGATGATGCTGCTTTGCTCCGATCTGAAGCGCGACGCTTC 1030
Db 772 CACACCCCGTCTTACTGTGATGATGCTGCTTTGCTCCGATCTGAAGCGCGCGACGCTTC 831
QY 1031 GGGGCGCGCGGATGCGAAGGGGAGTGCAGAAAGCTCGGTCTCTTGTGGAGCGCA 1090
Db 832 GGGGCGCGCGGATGCGAAGGGGAGTGCAGAAAGCTCGGTCTCTTGTGGAGCGCA 891


```

Db      649 ACAGACCATCAGCACAACCCCGCTTAAGTGTGATGATGCTGTGGCTCCCGGATCTG 708
Qy      1017 AAGCGCCAGCTTTGGGGGCCCCCGGATGCGAAAGGGGATGCCGAAAGTCCGTGCTC 1076
Db      709 AAGCGCCAGCTTTGGGGGCCCCCGGATGCGAAAGGGGATGCCGAAAGTCCGTGCTC 768
Qy      1077 TCTTGTGAGCGCCACCACTTAACCCCGGGGTACCACTGCGCCCGCCGACAGGGAGGAC 1136
Db      769 TCTTGTGAGCGCCACCACTTAACCCCGGGGTACCACTGCGCCCGCCGACAGGGAGGAC 828
Qy      1137 CACTGCGAGCCATTCAACCCAGCTGTCTCCCAAGACAGACCGGAGCCAGCGCAGACT 1196
Db      829 CACTGCGAGCCATTCAACCCAGCTGTCTCCCAAGACAGACCGGAGCCAGCGCAGACT 888
Qy      1197 CCCGAAAGTTATCTATGGGTGGGTAACTATCTTCAACCCCGCTTAACCGGGAGATTTCCTC 1256
Db      889 CCCGAAAGTTATCTATGGGTGGGTAACTATCTTCAACCCCGCTTAACCGGGAGATTTCCTC 948
Qy      1257 ATTCCCTCACCCACCTGAATGTCCCGACAGCTGTGATCTCCAGCCAGCAGTGGCCCCCACA 1316
Db      949 ATTCCCTCACCCACCTGAATGTCCCGACAGCTGTGATCTCCAGCCAGCAGTGGCCCCCACA 1008
Qy      1317 CAATGGGCTCAGAGATCTCAGGCTTTCACGTCAGGGGGTGAACCTTCGGAAGCCCTG 1376
Db      1009 CAATGGGCTCAGAGATCTCAGGCTTTCACGTCAGGGGGTGAACCTTCGGAAGCCCTG 1068
Qy      1377 TGACTTCAGAGGGCTCGCTCAGTTAGCATTTCAAGTGCATCAACCCCAAGAGTGGGG 1436
Db      1069 TGACTTCAGAGGGCTCGCTCAGTTAGCATTTCAAGTGCATCAACCCCAAGAGTGGGG 1128
Qy      1437 CCAACCCGAAACCAAAACCCCTTGAATGATCTCTTTCATCAGAGGTGGCTATGGGG 1496
Db      1129 CCAACCCGAAACCAAAACCCCTTGAATGATCTCTTTCATCAGAGGTGGCTATGGGG 1188
Qy      1497 CCAAGCGGAGAGATGAGCCCTTTCAGGGTAGAGAGACATTCATCAACCGGAAACC 1556
Db      1189 CCAAGCGGAGAGATGAGCCCTTTCAGGGTAGAGAGACATTCATCAACCGGAAACC 1248
Qy      1557 CAGGATTAAGAAGAGCCCTGAGGGGAGACAGACATAGAGGGGTGGAGAGTCTCC 1616
Db      1249 CAGGATTAAGAAGAGCCCTGAGGGGAGACAGACATAGAGGGGTGGAGAGTCTCC 1308
Qy      1617 CTTATCTGACATCTCTAGTCAATCTTGCCTTTCTCCGATTCGGAATTCGAGTGGGG 1676
Db      1309 CTTATCTGACATCTCTAGTCAATCTTGCCTTTCTCCGATTCGGAATTCGAGTGGGG 1368
Qy      1677 CCACTTAAAGATGCTCTCTCCAGCCCTGTCTCAACCATATCTCAAAATTAAGTCCAAC 1736
Db      1369 CCACTTAAAGATGCTCTCTCCAGCCCTGTCTCAACCATATCTCAAAATTAAGTCCAAC 1428
Qy      1737 CAGGGGCTGAGCACTCCCATCATCTTGTCTGCGCAAGTGGCAATTAAGGGGG 1796
Db      1429 CAGGGGCTGAGCACTCCCATCATCTTGTCTGCGCAAGTGGCAATTAAGGGGG 1488

RESULT 11
CR622791 1467 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CS0DL005Y024 of B cells (Ramos cell line)
DEFINITION full-length cDNA clone CS0DL005Y024 of B cells (Ramos cell line)
ACCESSION CR622791
VERSION CR622791.1 GI:50503598
KEYWORDS HTC; cDNA; full-length cDNA
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1467)
AUTHORS Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :

```

```

REFERENCE
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT - Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source
1.1467
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL005Y024"
/tissue_type="B cells (Ramos cell line) Cot 25-normalized"
/location="pCMVSPORT_6"

ORIGIN
Query Match 73.2%; Score 1325; DB 3; Length 1467;
Best Local Similarity 91.7%; Pred. No. 0;
Matches 1467; Conservative 0; Mismatches 0; Indels 132; Gaps 1;

Qy      177 TCCCTGCTGCTGGGTGCAAAAGTGTGGGTTCGTGGATTCGCGGGCCGTTTCAC 236
Db      1 TCCCTGCTGCTGGGTGCAAAAGTGTGGGTTCGTGGATTCGCGGGCCGTTTCAC 60
Qy      237 ACGTAGCTGTGCGGGCTCTCGGGTGAAGTCCGTCCGCGCGCGGTGCCCGGAGCGGCT 296
Db      61 ACGTAGCTGTGCGGGCTCTCGGGTGAAGTCCGTCCGCGCGCGGTGCCCGGAGCGGCT 120
Qy      297 AGGCTGCGGGGGTCCGGGGCCCGAGGCATTCGGGCTGCAATTGACGGGATCCCGGA 356
Db      121 AGGCTGCGGGGGTCCGGGGCCCGAGGCATTCGGGCTGCAATTGACGGGATCCCGGA 180
Qy      357 TGACACCGCGCGCCCGCCCGCCCTCAACGAGAGGTCCAGACCTGTGGGAAGAGTGGCG 416
Db      181 TGACACCGCGCGCCCGCCCGCCCTCAACGAGAGGTCCAGACCTGTGGGAAGAGTGGCG 240
Qy      417 GGAAGGGTCCCTGAGATCCCGATGCTAGCGCAAGATGCTCACTTATAGTGTGA 476
Db      241 GGAAGGGTCCCTGAGATCCCGATGCTAGCGCAAGATGCTCACTTATAGTGTGA 284
Qy      477 CTTACACATGTGACTTCACTCACTTGTGTGATCCGTAAATGACAAATTCGAGCTAC 536
Db      285 ----- 284
Qy      537 TTACAGTGTGTTGAGAGGATTAATGAACAATGCTTTAAAGCTTTTGACAGAGGG 596
Db      285 -----AGGG 288
Qy      597 AGCCTCGAAGACGAGGCTGTGCGCGAGACACCTGCTGTACCAAGGAGACCAAGGCA 656
Db      289 AGCCTCGAAGACGAGGCTGTGCGCGAGACACCTGCTGTACCAAGGAGACCAAGGCA 348
Qy      657 GCATGAAGACCCCGTGAAGCTGGCGGTCAATGATGAGACCTTCGAGTCCAGTACC 716
Db      349 GCATGAAGACCCCGTGAAGCTGGCGGTCAATGATGAGACCTTCGAGTCCAGTACC 408
Qy      717 GCTGCGAGGTGGCTACCGGGTCAAGGCAAGCGCATATGAGTGAAGTCTGTGTTG 776
Db      409 GCTGCGAGGTGGCTACCGGGTCAAGGCAAGCGCATATGAGTGAAGTCTGTGTTG 468
Qy      777 GCAAGCCAGAGGCAACCGGCTTACCCCAACCGGACTTCATCGCCCTGAGTGAAGG 836
Db      469 GCAAGCCAGAGGCAACCGGCTTACCCCAACCGGACTTCATCGCCCTGAGTGAAGG 528
Qy      837 CTACACGAACCAAGGCGTGGGCAAGAGGACATCGAAGGCTTTGGGGGCAAGGGAGCT 896

```

Db	529	CTAAACAGAACCAAGAGCGTGTGGGCAAGAGGAGTGGAAAGCCCTTGGGGGCAAAAGGGAGCT	588
Qy	897	GTGAGACCAACCCCTCTCAAGGGGCAAGCAACCCCACTCTCAACCAAGGAAGAACAAT	956
Db	589	GTGAGACCAACCCCTCTCAAGGGGCAAGCAACCCCACTCTCAACCAAGGAAGAACAAT	648
Qy	957	ACAGACCCATCAGGCACACCCCGTCTTACTGTGATGATGCGCTGTTTGGCTCCGATCTG	1016
Db	649	ACAGACCCATCAGGCACACCCCGTCTTACTGTGATGATGCGCTGTTTGGCTCCGATCTG	708
Qy	1017	AAGGCGCCAGCTTGGGGGCCCCGGAGATGCGCAAGGGGGATGCGCAAAAGCTCCGTCTC	1076
Db	709	AAGGCGCCAGCTTGGGGGCCCCGGAGATGCGCAAGGGGGATGCGCAAAAGCTCCGTCTC	768
Qy	1077	TCTTTGGAGCCCAACCACTTACCCCGAGGGGTAGGCATCTGCCCCGCCCCAGGAGGCAC	1136
Db	769	TCTTTGGAGCCCAACCACTTACCCCGAGGGGTAGGCATCTGCCCCGCCCCAGGAGGCAC	828
Qy	1137	CACGTGGAGCCATTCACCCAGCTGTGCTCCCTCAAGACAGACCCGGGGCAAGCT	1196
Db	829	CACGTGGAGCCATTCACCCAGCTGTGCTCCCTCAAGACAGACCCGGGGCAAGCT	888
Qy	1197	CCCAAGATTTCTATGGGTGGGTTACACTCTTCAACGCCCCCTGAAGGGGGGACTTTTCC	1256
Db	889	CCCAAGATTTCTATGGGTGGGTTACACTCTTCAACGCCCCCTGAAGGGGGGACTTTTCC	948
Qy	1257	ATTCCCTCACCCACCTGAATGTCCCGACACTGTGTATCGACCAACAGTGCCTCCCA	1316
Db	949	ATTCCCTCACCCACCTGAATGTCCCGACACTGTGTATCGACCAACAGTGCCTCCCA	1008
Qy	1317	CAATTTGGGCTCTCAGGATTTCTCAGGCTTTCACAGTCAAGGGGTGACTTTCGGAGCCCCCTGG	1376
Db	1009	CAATTTGGGCTCTCAGGATTTCTCAGGCTTTCACAGTCAAGGGGTGACTTTCGGAGCCCCCTGG	1068
Qy	1377	TGACTTCCAGGGCTCGCTCAGTTAGCAATTCAGTGCCATCTTACCACAAGAGGTGGGG	1436
Db	1069	TGACTTCCAGGGCTCGCTCAGTTAGCAATTCAGTGGCCATCTTACCACAAGAGGTGGGG	1128
Qy	1437	CCACCCAGAAACCAAAAGCCCCCTTGGAAATGATCTCTTTCATCAAGGTTGCTTATGGGG	1496
Db	1129	CCACCCAGAAACCAAAAGCCCCCTTGGAAATGATCTCTTTCATCAAGGTTGCTTATGGGG	1188
Qy	1497	CCACGGCGACAGGTATGGCCCCCTTGGCAGGGGTAGAGGACATTATCACCCAGGAAACC	1556
Db	1189	CCACGGCGACAGGTATGGCCCCCTTGGCAGGGGTAGAGGACATTATCACCCAGGAAACC	1248
Qy	1557	CAGGATTTAAAGAAAGCCCCGTGGGGGGGACAGACATAGGAGGGGTGGGCAATGGCTCC	1616
Db	1249	CAGGATTTAAAGAAAGCCCCGTGGGGGGGACAGACATAGGAGGGGTGGGCAATGGCTCC	1308
Qy	1617	CTTTATCTCGAACAAATCTCTAGTCGATTTCTCCCTTTTCTCCGATTTGGGAGTTGGGGG	1676
Db	1309	CTTTATCTCGAACAAATCTCTAGTCGATTTCTCCCTTTTCTCCGATTTGGGAGTTGGGGG	1368
Qy	1677	CCACCTCTAAGATGCTCTCTCCACGCCCCCTGTCTCAACCAATCTCAAAATTAGTCCAAAC	1736
Db	1369	CCACCTCTAAGATGCTCTCTCCACGCCCCCTGTCTCAACCAATCTCAAAATTAGTCCAAAC	1428
Qy	1737	CAGGGGCTGGCACTCCCAATATCATATGTCTTGCT	1775
Db	1429	CAGGGGCTGGCACTCCCAATATCATATGTCTTGCT	1467

RESULT	12
CR605763	
LOCUS	CR605763
DEFINITION	full-length cDNA clone CS0PC020YC08 of Neuroblastoma Cot
ACCESSION	CR605763
VERSION	CR605763.1 GI:50486570
KEYWORDS	HTC; CNSLT_cDNA.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
AUTHORS	1 (bases 1 to 1465)			
TITLE	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.			
JOURNAL	Full-length cDNA libraries and normalization unpublished			
REMARK	Contact : Feng Liang Email : fliang@lifeitech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600			
REFERENCE	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr			
AUTHORS	- Web : www.genoscope.cns.fr			
TITLE	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primers end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.			
COMMENT	Location/Qualifiers			
FEATURES	1..1465			
SOURCE	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="CSODC020YC08"			
	/tissue_type="Neuroblastoma Cot 25-normalized"			
	/plasmid="pCMVSPORT_6"			
ORIGIN				
Query Match	73.1%	Score 1323;	DB 3;	Length 1465;
Best Local Similarity	91.7%;	Pred. No. 0;		
Matches 1465;	Conservative 0;	Mismatches 0;	Indels 132;	Gaps 1
OY	179	CCTGGTTCCTGGGAGCAAGTGGCTGGGTTCTGGGTTTGTGATTGCGGGCGTTCAAC	238	
Db	1	CCTGGTTCCTGGGAGCAAGTGGCTGGGTTTCTGGGTTTGTGATTGCGGGCGTTCAAC	60	
OY	239	GTAGCTGTGTGCGGCTCCTCGGAGTACGTCGTCGCGCGCGAGTGCCTCGGAGCGGCT	298	
Db	61	GTAGCTGTGTGCGGCTCCTCGGAGTACGTCGTCGCGCGCGAGTGCCTCGGAGCGGCT	120	
OY	299	GCTGCGGGGGTCCGGGGCCCGCAGGCATTCCGGGCTGCAGATTGACGGGGATCCGGATG	358	
Db	121	GCTGCGGGGGTCCGGGGCCCGCAGGCATTCCGGGCTGCAGATTGACGGGGATCCGGATG	180	
OY	359	CACGCGCGGCGCGCGCGCGCTCACCGACGGTCCAGACTGTGGGAGAGAGTCCGGGG	418	
Db	181	CACGCGCGGCGCGCGCGCGCTCACCGACGGTCCAGACTGTGGGAGAGAGTCCGGGG	240	
OY	419	ACGGGTCCCTAGAGATCCCGATGCTACGACCAAGATGCTCAGCTTATAGTGTGACC	478	
Db	241	ACGGGTCCCTAGAGATCCCGATGCTACGACCAAGATGCTC-----	282	
OY	479	TACACATGTGACTTCACTCAGTTTGTGTGATCGTAAATGAGACAAATTGAAAGCTACTT	538	
Db	283	-----	282	
OY	539	CACAGTGTGTTGAGAGAGATTAAATGAACAATGCTGTAAAGCTCTTTGACAGAGGAG	598	
Db	283	-----AGGAG	288	
OY	599	CCTCGGAAGACGGGCTTGCGCGGAGACACACTGCTGTCAACGAGGAGACACAGGACAGC	658	
Db	289	CCTCGGAAGACGGGCTTGCGCGGAGACACACTGCTGTCAACGAGGAGACACAGGACAGC	348	
OY	659	ATGAAGACCCCGTGAAGCTGGCCGTCAGTGGATGACACCTTCGCGCTTCAAGACCGC	718	
Db	349	ATGAAGACCCCGTGAAGCTGGCCGTCAGTGGATGACACCTTCGCGCTTCAAGACCGC	408	
OY	719	TGCCGAGTGGCTACCGGGTCAAGGCCAGACGTCATATGTGATGAGACTCTGTTTGGC	778	
Db	409	TGCCGAGTGGCTACCGGGTCAAGGCCAGACGTCATATGTGATGAGACTCTGTTTGGC	468	

QY 1242 AGCGGGGACTTTCCCTCCCTCACTCCAGCACTGGAATGTCCTCCAGCACTGGTCACTCCAGCA 1301
 Db 481 AGCGGGGACTTTCCCTCCCTCACTCCAGCACTGGAATGTCCTCCAGCACTGGTCACTCCAGCA 540
 QY 1302 CCAAGTGGCCCAACAATAAGGGCTCAGAGATTCAGAGCTTCCAGTCAAGGGGTGACT 1361
 Db 541 CCAGTGGCCCAACAATAAGGGCTCAGAGATTCAGAGCTTCCAGTCAAGGGGTGACT 600
 QY 1362 TCCGAGGCCCCCTGGTGACTTCCAGAGGCTCCGCTCAGTGAATTCAGTTCAGAGCACTTACCC 1421
 Db 601 TCCGAGGCCCCCTGGTGACTTCCAGAGGCTCCGCTCAGTGAATTCAGTTCAGAGCACTTACCC 660
 QY 1422 CACGACGAGTGGGGCCACCCAGAAACCAAGCCCCCTTGGAAATGATCTTTCATCA 1481
 Db 661 CACGACGAGTGGGGCCACCCAGAAACCAAGCCCCCTTGGAAATGATCTTTCATCA 720
 QY 1482 GGGTTGCTTATGGGGCCACCGGACAGATGAGCCCTTTGCCAGGGTGAAGAGCAATTC 1541
 Db 721 GGGTTGCTTATGGGGCCACCGGACAGATGAGCCCTTTGCCAGGGTGAAGAGCAATTC 780
 QY 1542 TCACCCAGGAAACCCAGGATTAATAAGAGCCCTGTGGGGGACAGACATACAGAG 1601
 Db 781 TCACCCAGGAAACCCAGGATTAATAAGAGCCCTGTGGGGGACAGACATACAGAG 840
 QY 1602 GTGGGCACTGCTCCTTATCTGACATCTCTAGTCAATCTTCTGCTTTTCTCCGA 1661
 Db 841 GTGGGCACTGCTCCTTATCTGACATCTCTAGTCAATCTTCTGCTTTTCTCCGA 900
 QY 1662 TTGGGATTTGGGGGCACTCTTAAGATGCTCTCTCAAGCCCTGTCAACCATATCTCC 1721
 Db 901 TTGGGATTTGGGGGCACTCTTAAGATGCTCTCTCAAGCCCTGTCAACCATATCTCC 960
 QY 1722 AAATTAAGGCAACCCAGGCTGGGCACTTCCCAATCATCATCTTCTGCTCCAG 1781
 Db 961 AAATTAAGGCAACCCAGGCTGGGCACTTCCCAATCATCATCTTCTGCTCCAG 1020
 QY 1782 TGGCAATTAACGGCGTGAATGCCAA 1806
 Db 1021 TGGCAATTAACGGCGTGAATGCCAA 1045

RESULT 14
 BX358473
 LOCUS 1098 bp mRNA linear EST 08-APR-2004
 DEFINITION BX358473 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA
 clone CS0D1039YB15 5-PRIME, mRNA sequence.
 ACCESSION BX358473
 VERSION BX358473.2 GI:46283990
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1098)
 Li, W.B., Gruber, C., Jesssee, J. and Polayars, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 On May 2, 2003 this sequence version replaced gi:30347985.
 COMMENT
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 6579.r

For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?CS0D1039CA08QPl&c=6579.r.
 FEATURES
 source
 1. .1098
 Location/Qualifiers
 /organism="Homo sapiens"

/mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0D1039YB15"
 /cissue_type="PLACENTA COT 25-NORMALIZED"
 /clone_1ib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

Query Match 55.9%; Score 1012.2; DB 5; Length 1098;
 Best Local Similarity 98.4%; Pred. No. 1.7e-247;
 Matches 1060; Conservative 4; Mismatches 8; Indels 5; Gaps 4;

QY 196 AAGTCTGAGTTCTGGGTTTCTGATTCGCGGGCCGTTCAACGATGCTGTGCGGCTC 255
 Db 18 AAGTCTGAGTTCTGGGTTTCTGATTCGCGGGCCGTTCAACGATGCTGTGCGGCTC 77
 QY 256 CTGGGATGATCCGTCGCGGCGGGTGCCTCCGGGACGGCTTAAAGGCTGCGGGGTCCGGG 315
 Db 78 CTGGGATGATCCGTCGCGGCGGGTGCCTCCGGGACGGCTTAAAGGCTGCGGGGTCCGGG 137
 QY 316 GCCCAGGCAATTCGCGGCTGAGATTCAGCGGGATCCCGATGACCGCGCCCGCG 375
 Db 138 GCCCAGGCAATTCGCGGCTGAGATTCAGCGGGATCCCGATGACCGCGCGCCCGCG 196
 QY 376 CCTTCACGAGCGGTCAGACCTGTGGGAAAGAGTCCGGGACGGGTCCCTGAGATC 435
 Db 197 CCTTCACGAGCGGTCAGACCTGTGGGAAAGAGTCCGGGACGGGTCCCTGAGATC 256
 QY 436 CCGATCCCTAGAGCGCAAGATGCTCAAGTTATAGTGTACCTACATAGTGAATTCAC 495
 Db 257 CCGATCCCTAGAGCGCAAGATGCTCAAGTTATAGTGTACCTACATAGTGAATTCAC 316
 QY 496 CTCAGTTTGTATCCGTAATGGAACAATTCGAACTACTTCACTGCTGTGAGAG 555
 Db 317 CTCAGTTTGTATCCGTAATGGAACAATTCGAACTACTTCACTGCTGTGAGAG 376
 QY 616 GCGCGGACAGACACTGCTGTCAACGAGGACCAAGCATGAAGACCCCGTGA 675
 Db 437 GCGCGGACAGACACTGCTGTCAACGAGGACCAAGCATGAAGACCCCGTGA 496
 QY 676 GCTGGCGGTGAGTGGATGACAGACCTGCGCTTCAGACCGCTGCGAGGTGCTACCG 735
 Db 497 GCTGGCGGTGAGTGGATGACAGACCTGCGCTTCAGACCGCTGCGAGGTGCTACCG 556
 QY 736 GGTCAAGGCAAGACGTATATGTGATGAGACTCTGTTTGGACAGCCAGGACCGCG 795
 Db 557 GGTCAAGGCAAGACGTATATGTGATGAGACTCTGTTTGGACAGCCAGGACCGCG 616
 QY 796 GCTTACCCACCGGACTTTCATCCGCTGCTGGTGGAGAAAGCTTAACAGAACAGGCGT 855
 Db 617 GCTTACCCACCGGACTTTCATCCGCTGCTGGTGGAGAAAGCTTAACAGAACAGGCGT 676
 QY 856 GGGCAAGGACATGAAAGGCTTGGGGGCAAGGGGAGCTGTGAGACACCCCTCAAG 915
 Db 677 GGGCAAGGACATGAAAGGCTTGGGGGCAAGGGGAGCTGTGAGACACCCCTCAAG 736
 QY 916 GGGGAGCAACCCCAACCTTCACACCAAGGAAAGAAACAATATACAGACCATCAGCACAC 975
 Db 737 GGGGAGCAACCCCAACCTTCACACCAAGGAAAGAAACAATATACAGACCATCAGCACAC 796
 QY 976 CCCGTTTACTGTATGATGAGTGTGTTGGTCCCGATCTGAAGCGGCACTTCTGGGGC 1035
 Db 797 CCCGTTTACTGTATGATGAGTGTGTTGGTCCCGATCTGAAGCGGCACTTCTGGGGC 856
 QY 1036 CCCGCGATGCGAAGGGGATGCGGCAAGCTCCGATCTCTTGTGAGCGCCACGAC 1095

Db 857 CCGCGGATGGCGAAGGGGATG-CGCAAGCTCGTCTCTTGTGAGCGCCACC 915
 QY 1096 TACCCCGAGGGGTGACCACTC--GCCCCCGCCAGGGAGGACCACTGCGGAGCATTCAC 1153
 Db 916 TACCCCGAGGGGTGACCACTC--GCCCCCGCCAGGGAGGACCACTGCGGAGCATTCAC 975
 QY 1154 CCGAGCTGTCCTCCCAAGAGAGAGCGGGGCGAGCGGACACTCCAGAAATTATCATG 1213
 Db 976 CCGAGCTGTCCTCCCAAGAGAGAGCGGGGCGAGCGGACACTCCAGAAATTATCATG 1035
 QY 1214 GGTGGGTAACTCTTCAAGCGCCCTGTAAGCGGGGACTTCCATTCCTTCACCCAC 1270
 Db 1036 GGTGGGTAACTCTTCAAGCGCCCTGTAAGCGGGGACTTCCATTCCTTCACCCAC 1091
 RESULT 15
 BX339727 1038 bp mRNA linear EST 07-APR-2004
 LOCUS BX339727 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
 DEFINITION clone CSOD1076YG21 5-PRIME, mRNA sequence.
 ACCESSION BX339727 GI:46271262
 VERSION BX339727.2
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 1038)
 AUTHORS Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT On May 2, 2003 this sequence version replaced gi:30335832.
 CONTACT: Genoscope - Centre National de Sequencage
 Genoscope - Centre National de Sequencage
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 6579.r
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdnare-CSOD1076YAD11QPlsc-6579.r.
 FEATURES
 SOURCE
 Location/Qualifiers
 1..1038
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CSOD1076YG21"
 /issue_type="PLACENTA COT 25-NORMALIZED"
 /clone_id="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo (dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."
 ORIGIN
 Query Match 53.9%; Score 976.8; DB 5; Length 1038;
 Best Local Similarity 99.0%; Pred. No. 1.8e-238;
 Matches 1008; Conservative 5; Mismatches 2; Indels 3; Gaps 3;
 QY 762 ATGAGACTCTGTTGGAGCCCAAGGACCCCGCTACCCCAACCGAATTGATCCG 821
 Db 1 ATGAGACTCTGTTGGAGCCCAAGGACCCCGCTACCCCAACCGAATTGATCCG 60
 QY 822 CTTGGGTGAGAGGCTTAACAGAACAGAGGCGTGGCAAGAGGATCGAAGGCTTGG 881
 Db 61 CTTGGGTGAGAGGCTTAACAGAACAGAGGCGTGGCAAGAGGATCGAAGGCTTGG 120
 QY 882 GGGCAAGAGGAGCTGTGAGACCAACCCCTCAAGGGGAGCAACCCCTCAACCA 941
 Db 121 GGGCAAGAGGAGCTGTGAGACCAACCCCTCAAGGGGAGCAACCCCTCAACCA 180

QY 942 GGAAGAGAACAATACAGACCCATGAGCCACACCCCGCTTACTGTGATGATCCCTGT 1001
 Db 181 GGAAGAGAACAATACAGACCCATGAGCCACACCCCGCTTACTGTGATGATCCCTGT 240
 QY 1002 TTGGCTCCGATGTGAAGGCGCAGCTTGGGGGCCCCGGATGCGAAGGGGATGCCG 1061
 Db 241 TTGGCTCCGATGTGAAGGCGCAGCTTGGGGGCCCCGGATGCGAAGGGGATGCCG 300
 QY 1062 CAAGCTCCGCTGCTCTTGTGAGACCCCACTACCTACCCCAAGGGGATGCCAATCGGCC 1121
 Db 301 CAAGCTCCGCTGCTCTTGTGAGACCCCACTACCTACCCCAAGGGGATGCCAATCGGCC 360
 QY 1122 GCCCAAGAGAGGACACTGCGAGCCATTACCACTGCTCCCTCAAGAGAGAGCCGG 1181
 Db 361 GCCCAAGAGAGGACACTGCGAGCCATTACCACTGCTCCCTCAAGAGAGAGCCGG 420
 QY 1182 GGCACAGGAGACTCCAGAGAGTTATCTATGGGTGGTTACACTCTTCACGCCCTGTA 1241
 Db 421 GGCACAGGAGACTCCAGAGAGTTATCTATGGGTGGTTACACTCTTCACGCCCTGTA 480
 QY 1242 AGCGGGGACTTTCCCATTCCTTCACCACTGATGTCCTGACACTGGTCATCCAGCCA 1301
 Db 481 AGCGGGGACTTTCCCATTCCTTCACCACTGATGTCCTGACACTGGTCATCCAGCCA 540
 QY 1302 CCAAGTCCCCCAACAAATGGGCTCAGGATCTCAGGCTTCCAGGTCAGGGGTGACT 1361
 Db 541 CCAAGTCCCCCAACAAATGGGCTCAGGATCTCAGGCTTCCAGGTCAGGGGTGACT 600
 QY 1362 TCCGAGACCCCTGTGTGACTTCCAGAGGCTGCTCAGTATTCAGTTCAGTTCAGCC 1421
 Db 601 TCCGAGACCCCTGTGTGACTTCCAGAGGCTGCTCAGTATTCAGTTCAGTTCAGCC 660
 QY 1422 CACGACGAGTGGGGCCACCCAGAAACCAAGCCCTTGGAAATGATCTTTTCATCA 1481
 Db 661 CACGACGAGTGGGGCCACCCAGAAACCAAGCCCTTGGAAATGATCTTTTCATCA 720
 QY 1482 GGGTTGCTATGGGGCCACCGGCGACAGTATGGCCCTTCCAGAGGTAGAGACATTC 1541
 Db 721 GGGTTGCTATGGGGCCACCGGCGACAGTATGGCCCTTCCAGAGGTAGAGACATTC 780
 QY 1542 TCACCCAGGAAACCCAGATTTAAGAGACCCCTGTGGGGGACAGACATAGCAGGG 1601
 Db 781 TCACCCAGGAAACCCAGATTTAAGAGACCCCTGTGGGGGACAGACATAGCAGGG 839
 QY 1602 GTGGGAGTGGCTCCCTTATCCGACATCTAGTCAATCTGCTGCTTTTCTCCGA 1661
 Db 840 GTGGGAGTGGCTCCCTTATCCGACATCTAGTCAATCTGCTGCTTTTCTCCGA 899
 QY 1662 TTGGGATTTGGGGCCACCTCTTAAGATGCTCTTCAGCCCTGTCTCAACATATCC 1721
 Db 900 TTGGGATTTGGGGCCACCTCTTAAGATGCTCTTCAGCCCTGTCTCAACATATCC 959
 QY 1722 AATTTAGTCCAAACCCAGGGGCTGTGCACTCCCAATTCATTTGCTGCGCA 1779
 Db 960 AATTTAGTCCAAACCCAGGGGCTGTGCACTCCCAATTTGCTGCGCA 1015

Search completed: March 28, 2005, 05:53:49
 Job time : 6168 secs

Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.

This Page Blank (uspto)

BEST AVAILABLE COPY

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2005, 05:53:55 ; Search time 80 Seconds
(without alignments)
1300.482 Million cell updates/sec

Title: US-10-031-589-4
Perfect score: 1437
Sequence: 1 MKTPVELVSGMOTGLQHR.....SVSPRRRGATQKRPWK 269

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1437	100.0	269	4 AAB67330	Aab67330 Human neu
2	1437	100.0	269	4 AAB94540	Aab94540 Human pro
3	1216	84.6	229	8 ADO57487	Ado57487 Hairless
4	863	60.1	177	4 AAB94561	Aab94561 Human pro
5	239	16.6	467	4 ABB11341	Abb11341 Human typ
6	218	15.2	65	5 ABB35240	Abb35240 Human ORF
7	147.5	10.3	1706	7 ABO83014	Ab083014 Pseudomon
8	136.5	9.5	4019	4 AAE13839	Aae13839 Human lun
9	136.5	9.5	4019	7 ADB6733	Ad66733 Human lun
10	136.5	9.5	4019	7 ADE87987	Ad87987 Human lun
11	136.5	9.5	4025	5 ABB69736	Abb69736 Human pol
12	136	9.5	2703	4 ABB63299	Abb63299 Drosophi
13	136	9.5	2703	5 ABB63299	Abb63299 Drosophi
14	135.5	9.4	920	7 ABO80172	Ab080172 Pseudomon
15	134.5	9.4	19938	6 ABE76681	Ab76681 Streptomy
16	134	9.3	502	6 ADA24266	Ada24266 Human WAS
17	132.5	9.2	488	7 ABB21932	Abb21932 Novel hum
18	132.5	9.2	488	7 ADG73027	Adg73027 Pseudomon
19	132.5	9.2	488	7 ADL12082	Adl12082 Pseudomon
20	132.5	9.2	580	8 ADO19877	Ado19877 Human sof
21	132.5	9.2	1008	7 ABO80673	Ab080673 Pseudomon
22	131.5	9.2	705	4 AAM39328	Aam39328 Human pol
23	131.5	9.2	714	4 AAM41114	Aam41114 Human pol
24	129.5	9.0	791	5 ABB65089	Abb65089 Hypoxia-r
25	129.5	9.0	791	8 ADM03640	Adm03640 Antipsoi

26	129.5	9.0	791	8 ADP22972	Adp22972 PRO polyp
27	129.5	9.0	863	6 ABP98856	Abp98856 Human ser
28	129.5	9.0	863	7 ADC31071	Adc31071 Human nov
29	129.5	9.0	863	7 ADE48296	Ad48296 Human MIC
30	129.5	9.0	863	7 ADE48308	Ad48308 Human MIC
31	129.5	9.0	863	8 ABM82355	Abm82355 Tumour-a8
32	129	9.0	934	8 ADN38399	Adn38399 Novel hum
33	128.5	8.9	501	7 ABO68984	Ab068984 Pseudomon
34	128.5	8.9	536	2 ABO68008	Ab068008 Pseudomon
35	128	8.9	504	2 AA124091	Aay24091 Human Wls
36	128	8.9	566	6 ADA24286	Ada24286 Protein r
37	127.5	8.9	1870	7 ADJ68314	Adj68314 Human hea
38	126.5	8.8	693	7 ABO77325	Ab077325 Pseudomon
39	126.5	8.8	1199	7 ADD46005	Add46005 Rat Prote
40	126.5	8.8	1199	7 ADE61171	Ad61171 Rat Prote
41	126.5	8.8	1321	8 ADN98332	Adn98332 Mouse ins
42	126	8.8	718	5 ABB79534	Abb79534 Human tum
43	126	8.8	958	6 ABB75918	Abb75918 Human sec
44	125.5	8.7	641	7 ADG48274	Adg48274 Human ret
45	125.5	8.7	663	7 ADG48269	Adg48269 Human ret

ALIGNMENTS

RESULT 1
ID AAB67330 standard; protein; 269 AA.
XX
AC AAB67330;
XX
DT 23-APR-2001 (first entry)
XX
DE Human neuron progenitor cell clone #2 protein.
XX
KW Neuron; progenitor cell; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200107607-A2.
XX
PD 01-FEB-2001.
XX
PF 21-JUL-2000; 2000MO-JP004895.
XX
PR 23-JUL-1999; 99JP-00209817.
XX
PR 18-OCT-1999; 99US-0159528P.
XX
PA (HELI-) HELIX RES INST.
XX
Ota T, Isogai T, Nishikawa T, Kawai Y;
WPI, 2001-182791/18.
XX
PT New human polynucleotides, particularly DNAs, isolated from a cDNA
PT library derived from progenitor cells, useful in gene therapy, as well as
PT in producing proteins useful as diagnostic markers in drug development.
XX
PS Claim 1; Page 36-37; 54pp; English.
XX
CC The present invention relates to human proteins isolated from clones from
CC neuron progenitor cells. The proteins and the DNA encoding them may be
CC used in the preparation of treatments for diseases associated with the
CC proteins
SQ Sequence 269 AA;
Query Match 100.0%; Score 1437; DB 4; Length 269;
Best Local Similarity 100.0%; Pred. No. 2.7e-114;
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MKTPVELVSGMOTGLQHRGGRVAKRTSYDETLFGSPAGTRPPDPDPWVEKA 60
|||||

Db 1 MKTPVELAVSGMOTLGIQHRGCGYRVKARTSYVDLTFGSPAGTRPTPPDFPWEKA 60
 QY 61 NRTGVGKEASKALGAKGSCETTPSRGSTPTLTPRKKNKRPISHTPSYCDSEIFGSRSE 120
 Db 61 NRTGVGKEASKALGAKGSCETTPSRGSTPTLTPRKKNKRPISHTPSYCDSEIFGSRSE 120
 QY 121 GASFGAPRMKGDAAKLRALMLTPTPTPRGSHSPREAPLRAIHPAGSKTEPGPAADS 180
 Db 121 GASFGAPRMKGDAAKLRALMLTPTPTPRGSHSPREAPLRAIHPAGSKTEPGPAADS 180
 QY 181 OKLSMGLHSRPLKRGLSHTLNLVPSYGHATSPHTNGPDLRPSYGVTFRRSPLV 240
 Db 181 OKLSMGLHSRPLKRGLSHTLNLVPSYGHATSPHTNGPDLRPSYGVTFRRSPLV 240
 QY 241 TSRRASVSISVPTPRRGATQKPKPPWK 269
 Db 241 TSRRASVSISVPTPRRGATQKPKPPWK 269
 Db 241 TSRRASVSISVPTPRRGATQKPKPPWK 269
 RESULT 2
 AAB94540
 ID AAB94540 standard; protein; 269 AA.
 AC AAB94540;
 DT 26-JUN-2001 (first entry)
 DE Human protein sequence SEQ ID NO:15284.
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 OS Homo sapiens.
 PN EPI074617-A2.
 PD 07-FEB-2001.
 PF 28-JUL-2000; 2000EP-00116126.
 PR 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 PA (HELI-) HELIX RES INST.
 PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ihii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI; 2001-318749/34.
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.
 PS Claim 8; SEQ ID NO 15284; 2537bp + Sequence listing; English.
 The present invention describes primer sets for synthesizing 5602 full-
 length cDNAs defined in the specification. Where a primer set comprises:
 (a) an oligo-dT primer and an oligonucleotide complementary to the
 complementary strand of a polynucleotide which comprises one of the 5602
 nucleotide sequences defined in the specification, where the
 oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 of an oligonucleotide comprising a sequence complementary to the
 complementary strand of a polynucleotide which comprises a 5'-end
 sequence and an oligonucleotide comprising a sequence complementary to a
 polynucleotide which comprises a 3'-end sequence, where the
 oligonucleotide comprises at least 15 nucleotides and the combination of
 the 5'-end sequence/3'-end sequence is selected from those defined in the
 specification. The primer sets can be used in antisense therapy and in
 gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX Sequence 269 AA;
 SQ
 Query Match 100.0%; Score 1437; DB 4; Length 269;
 Best Local Similarity 100.0%; Pred. No. 2.7e-114;
 Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKTPVELAVSGMOTLGIQHRGCGYRVKARTSYVDLTFGSPAGTRPTPPDFPWEKA 60
 Db 1 MKTPVELAVSGMOTLGIQHRGCGYRVKARTSYVDLTFGSPAGTRPTPPDFPWEKA 60
 QY 61 NRTGVGKEASKALGAKGSCETTPSRGSTPTLTPRKKNKRPISHTPSYCDSEIFGSRSE 120
 Db 61 NRTGVGKEASKALGAKGSCETTPSRGSTPTLTPRKKNKRPISHTPSYCDSEIFGSRSE 120
 QY 121 GASFGAPRMKGDAAKLRALMLTPTPTPRGSHSPREAPLRAIHPAGSKTEPGPAADS 180
 Db 121 GASFGAPRMKGDAAKLRALMLTPTPTPRGSHSPREAPLRAIHPAGSKTEPGPAADS 180
 QY 181 OKLSMGLHSRPLKRGLSHTLNLVPSYGHATSPHTNGPDLRPSYGVTFRRSPLV 240
 Db 181 OKLSMGLHSRPLKRGLSHTLNLVPSYGHATSPHTNGPDLRPSYGVTFRRSPLV 240
 QY 241 TSRRASVSISVPTPRRGATQKPKPPWK 269
 Db 241 TSRRASVSISVPTPRRGATQKPKPPWK 269
 Db 241 TSRRASVSISVPTPRRGATQKPKPPWK 269
 RESULT 3
 AD057487
 ID AD057487 standard; protein; 229 AA.
 AC AD057487;
 DT 29-JUL-2004 (first entry)
 DE Hairless protein interaction partner #34.
 KW human; hairless protein; Hrt; ubiquitous receptor UR; MAP1A;
 KW KIAA0930 protein; monocytes antigen CD14; sphingolipid activator protein;
 KW beta-synuclein; C11 protein; vesicle-associated membrane protein 2;
 KW aldolase A; CGI-106 protein; hypothalamus protein HSMNP1; alpha enolase;
 KW POM-2P3; guanine oxidoreductase; pumilio 1; VPS41; KIAA0614 protein;
 KW splicing factor CCl.4; ubiquitin; beta-mannosidase; hair growth.
 OS Homo sapiens.
 PN US2004086945-A1.
 PD 06-MAY-2004.
 PR 02-JUN-2003; 2003US-00452858.
 PR 03-JUN-2002; 2002US-0385414P.
 PA (PROC) PROCTER & GAMBLE CO.
 PI Sreekrishna K, Gerwe GS, Toerner DR;
 DR WPI; 2004-430095/40.
 DR N-FSDB; AD057486.
 PT New composition comprising mouse truncated hairless protein-human
 PT interacting partner protein or nucleic acid complexes, useful for
 PT screening test compounds that inhibit or enhance hair growth.

XX Example 1; SEQ ID NO 79; 60pp; English.
 CC The invention relates to a composition comprising a mouse truncated
 CC hairless (Hr) protein-human interacting partner protein or nucleic acid
 CC complex. The human interacting partner protein comprises a molecule
 CC selected from ubiquitinous receptor UR, MAP1A, KIAA0930 protein, monocytes
 CC antigen CD14, sphingolipid activator protein, beta-synuclein, C11
 CC protein, vesicle-associated membrane protein 2, aldolase A, CGI-106
 CC protein, hypochalamin protein HSNP1, alpha enolase, POM-2P3, quinone
 CC oxidoreductase, pumilio 1, VPS41, KIAA0614 protein, splicing factor
 CC CCL1.4, ubiquitin, beta-mannosidase. Also described are: methods of
 CC assaying a test compound for agonist or antagonist activity for the above
 CC composition; and methods of inhibiting or increasing hair growth on a
 CC surface in a subject. The present sequence represents a hairless protein
 CC human interacting partner used in the composition.

Query Match 84.6%; Score 1216; DB 8; Length 229;
 Best Local Similarity 98.7%; Pred. No. 1.6e-95;
 Matches 226; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 11 GMDTGLQHRGCGYRVKARTSYVDLFGSPAGTRPPDPDPWVEKANTRGVKEA 70
 DB 1 GMDTGLQHRGCGYRVKARTSYVDLFGSPAGTRPPDPDPWVEKANTRGVKEA 60
 QY 71 SKALGAKSCETTPRGSTPTLTPKKNKYRPISTPSCDLSGSRSEGSFGAPRMA 130
 DB 61 SKALGAKSCETTPRGSTPTLTPKKNKYRPISTPSCDLSGSRSEGSFGAPRMA 120
 QY 131 KGDAKALRALMTPTTPRGSHSPRPREAPLAIHPAGSKTEPGPADSQGLSMGGLHS 190
 DB 121 KGDAKALRALMTPTTPRGSHSPRPREAPLAIHPAGSKTEPGPADSQGLSMGGLHS 180
 QY 191 SRPLRGKLSHSLTLNLPSTGHPATSPHTNGPQDLRPSTGVTSP 239
 DB 181 SRPLRGKLSHSLTLNLPSTGHPATSPHTNGPQDLRPSTGVTSP 229

RESULT 4
 AAB94561
 ID AAB94561 standard; protein; 177 AA.

XX AC AAB94561;

XX DT 26-JUN-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:15335.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

XX FN EPI074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-00116126.

XX PR 29-JUL-1999; 99JP-00248036.

XX PR 27-AUG-1999; 99JP-00300253.

XX PR 11-JAN-2000; 2000JP-00118776.

XX PR 02-MAY-2000; 2000JP-00183767.

XX PR 09-JUN-2000; 2000JP-00241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Iisogi T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX PI WPI, 2001-31879/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.

XX Claim 8; SEQ ID NO 15335; 2537bp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo- or primer and an oligonucleotide which comprises one of the 5602
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification; where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention

SEQ Sequence 177 AA;

Query Match 60.1%; Score 863; DB 4; Length 177;
 Best Local Similarity 99.4%; Pred. No. 1.5e-65;
 Matches 160; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTPELVSGMDTGLQHRGCGYRVKARTSYVDLFGSPAGTRPPDPDPWVEKA 60
 DB 1 MKTPELVSGMDTGLQHRGCGYRVKARTSYVDLFGSPAGTRPPDPDPWVEKA 60
 QY 61 NTRGVGEASKALGAKSCETTPRGSTPTLTPKKNKYRPISTPSCDLSGSRSE 120
 DB 61 NTRGVGEASKALGAKSCETTPRGSTPTLTPKKNKYRPISTPSCDLSGSRSE 120
 QY 121 GASFGAPPMAGDAKALRALMTPTTPRGSHSPRPREAPL 161
 DB 121 GASFGAPPMAGDAKALRALMTPTTPRGSHSPRPREAPL 161

RESULT 5

ABBI1341
 ID ABBI1341 standard; peptide; 467 AA.

XX AC ABBI1341;

XX DT 11-JAN-2002 (first entry)

XX DE Human type II procollagen homologue, SEQ ID NO:1711.

XX KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
 KW haematopoiesis regulation; tissue growth; immunomodulation; activin;
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
 KW chronic inflammatory condition; proliferative retinopathy;
 KW atherosclerosis; coronary heart disease; arterial ischaemia;
 KW bone disorder; osteoporosis; vascular growth disorder;
 KW tissue regeneration; wound healing; infection; immune disorder;
 KW cell culture; drug screening; gene therapy; anti-inflammatory;
 KW antiaesthetic; antiarthritis; haemostatic; antiarteriosclerotic;
 KW cyostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
 KW antifungal; vulnerary; antilucer.

OS Homo sapiens.
 XX W0200157188-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US003800.
 XX
 PR 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Dermanac RT;
 XX
 DR WPI; 2001-457740/49.
 DR N-PSDB; ABA08585.
 XX
 PT Human proteins and DNA encoding sequences useful for preventing, treating
 PT or ameliorating a medical condition in a mammalian subject e.g. arthritis
 PT and cancer.
 XX
 PS Claim 20; Page 167-168; 1963pp; English.
 XX
 CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
 CC invention also relates to vectors and recombinant host cells comprising a
 CC nucleotide of the invention, methods of producing the novel polypeptides,
 CC antibodies against the polypeptides, methods of detecting the nucleotides
 CC or polypeptides in a sample, and methods of identifying compounds which
 CC bind to polypeptides of the invention. Although novel, many of the
 CC polypeptides of the invention have homology to known proteins, thereby
 CC giving an insight into their probable biological activities, and hence
 CC potential therapeutic applications. The polypeptides of the invention may
 CC have various activities, including cytokine, cell proliferation or cell
 CC differentiation activities; stem cell growth factor activity;
 CC hematopoietic regulatory activity; tissue growth activity;
 CC immunomodulatory activity; activin- or inhibin-related activities;
 CC chemokine or chemokine activities; haemostatic, thrombotic or
 CC thrombolytic activities; receptor or ligand activities; or may be
 CC involved in oncogenesis, cancer cell proliferation or metastasis.
 CC Depending on their biological activities, polypeptides and nucleotides of
 CC the invention are useful for preventing, treating or ameliorating medical
 CC conditions, e.g., by protein or gene therapy. Such conditions include
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
 CC vascular growth. Polypeptides involved with tissue regeneration and
 CC repair (or nucleic acids encoding them) may be used to promote wound
 CC healing (e.g., of burns, incisions and ulcers), while those with
 CC immunomodulatory activities may be used in the treatment of viral,
 CC bacterial and fungal infections in addition to immune disorders.
 CC Polypeptides with growth factor activity may be used in cell cultures to
 CC promote cell growth. For example, such polypeptides may be used to
 CC manipulate stem cells in culture to give rise to neuroepithelial cells
 CC that can be used to augment or replace cells damaged by illness,
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides
 CC may also be used in the diagnosis of the above conditions, and in drug
 CC screening techniques. The present sequence represents a novel human
 CC polypeptide of the invention
 XX
 SO Sequence 467 AA;
 XX
 QY Query Match 16.6%; Score 239; DB 4; Length 467;
 Db Best Local Similarity 29.4%; Pred. No. 8.9e-12;
 Matches 94; Conservative 30; Mismatches 94; Indels 102; Gaps 17;
 QY 5 VELAVSGMQLTGLQHR--CRGGYR-----VKARTSVVDLTGLSGPACTPTPPDPDP 56
 Db 176 MRLCLAAQQAQPGALPHRTSIRPGWRLTEPAMAR--HRRPWGQRCVAVPPQGAAPP 232
 QY 57 VEKANRTGVGKESKALGAKGSCETTPRSGSTPTLTLPKKNKRPISHTPSYC---DES 113

Db 233 SHQGRT-----NTDPS--ATPRLT-----VMSGCLAPDLK 261
 QY 114 LFGSRSEAGSFGAPRNAKDAAKLRLALWTPPTPGSGSPRRAPLRAIHPAGSKTE 173
 Db 262 APASGPRGWRGMPQGS-----GALLWTPPTPGSGSPRRAPLRAIHPAGSKSR 314
 QY 174 PGPAADSQKLSMGLHSSRLKRG-----LSHSL-----THLVNPTGHPATS 216
 Db 315 AGASGRLEVTYGMVTLFTPEAGTFLDSFTYMSPALVIQPPVPTQGLRISGLPRQG 374
 QY 217 APHTNGPQDLRPSTSGVTF-----RSPLVTSRASVSISVSTPRR----- 257
 Db 375 XP-SGAPFWLI-PELAQLAFQCHLPHDEVGPRRQSLGND---TLSSGLPMGRQVWPL 429
 QY 258 ---GG--ATQKP---KPPW 268
 Db 430 ARVGHSSPREPQVLKKPLM 449
 RESULT 6
 ID ABB35240
 ABP35240 standard; protein; 65 AA.
 XX
 AC ABB35240;
 XX
 DT 08-JUL-2002 (first entry)
 XX
 DE Human ORP4213 protein, SEQ ID NO:8426.
 XX
 KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;
 KW immune modulation; haematopoiesis regulation; tissue growth;
 KW angiogenesis; activin; inhibin; chemotactic; chemokine; haemostatic;
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
 KW behaviour; cancer; proliferative disorder; neurological disorder;
 KW cardiovascular disease; immune system disorder; organ transplantation;
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;
 KW vasotropic; antipruritic; antidiabetic; cyostatic; nootropic;
 KW neuroprotective; antiatherosclerotic; anticoagulant; immunomodulator;
 KW cardiant; hypotensive; antihypertoid; antiinflammatory; immunomodulator;
 KW dermatological; analgesic; virucide; antibacterial; fungicide.
 OS Homo sapiens.
 XX
 PN W0200190366-A2.
 XX
 PD 29-NOV-2001.
 XX
 PF 24-MAY-2001; 2001WO-US017076.
 XX
 PR 24-MAY-2000; 2000US-020690P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Leach MD, Shinkets RA;
 XX
 DR WPI; 2002-106200/14.
 DR N-PSDB; ABB79266.
 XX
 PT Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and disorders related to organ
 PT transplantation.
 XX
 PS Claim 10; Page 2350; 2508pp; English.
 XX
 CC Sequences ABB1028-ABB35561 represent 4534 novel human proteins
 CC designated ORF (open reading frame) 1-4534, and sequences ABB75054-
 CC ABB79587 represent cDNAs encoding them. The invention also encompasses
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to

the ORFX nucleic acid sequences, vectors and host cells comprising ORFX polynucleotides, the recombinant production of ORFX proteins, antibodies specific for ORFX proteins, methods of detecting ORFX polynucleotides and polypeptides, methods of screening for modulators of ORFX expression or activity, and methods of screening individuals for a predisposition to an ORFX-associated disorder. The ORFX proteins of the invention have a wide range of biological activities, such as cytokine, cell proliferation, cell differentiation, immune modulation, haematopoiesis regulation, cell growth, angiogenesis, activin or inhibin activity, chemotactic/chemokinetic activity, haemostatic activity, thrombolytic activity, receptor/ligand, anti-inflammatory activity, tumour inhibition activity, and antineoplastic activity, and may also be involved in the determination of bodily characteristics, fertility and behaviour. ORFX proteins, nucleic acids and antibodies may be used in the treatment of cancers, other proliferative disorders such as psoriasis and benign tumours, neurological disorders such as epilepsy and Alzheimer's disease, cardiovascular diseases, immune system disorders, disorders related to organ transplantation, disorders of tissue growth and regeneration, diseases such as diabetes mellitus, hypothyroidism, and cholesterol, ester, sugar and other pathogens. ORFX nucleic acids may also be used as a source of primers and probes, in the detection of ORFX genomic sequences or transcripts, in the identification and cloning of homologous sequences, in genetic diagnosis, and in forensic biology. The ORFX nucleic acids may additionally be used to produce transgenic animals which may be useful for studying the function and/or activity of ORFX protein, and in drug screening. The ORFX proteins may also be used as immunogens to generate specific antibodies, which are useful in the diagnosis, treatment and monitoring of ORFX-associated diseases

Query Match	15.2%	Score 218	DB 5	Length 65
Best Local Similarity	71.4%	Pred. No. 5e-11		
Matches 40; Conservative	4	Mismatches 12	Indels 0	Gaps 0

```

QY      12  MQTLLGQHRCRGGYRVKARTSYVDETLGSPAGTRTPPDFFDPWVEKAKARTRGVG 67
      | : : ||||| ||||| ||||| ||||| ||||| : ||: ||||
Db      1  MKAIIHQRSPTSYRVKARASVYDETLGSPARTRAQPDFDPWVQNCNRSRGVG 56

```

RESULT 7	
ABO83014	
ID ABO83014	standard; protein; 1706 AA

DT 29-JUL-2004 (first entry)

DE Pseudomonas aeruginosa polypeptide #15189

KW Bacterial infection; *Pseudomonas aeruginosa* infection; antibacterial ...

05 Pseudomonas aeruginosa.

PN US6551795-B1.

PD 22-APR-2003

PF 18-FEB-1999; 99US-00252991.

PR 18-FEB-1998; 98US-0074788P

XX
FZ

(CENO) REMOVE FROM SECRET AND DESTROY

XX
DT Pubhoefojd NT 80771-6 T 279-1-1-1-1 C 2-1-1-1 D

XX WPT 2003 215200/50
DE

DR N-PSDB; ABD16585.
yy

PT Novel isolated nucleic acid encoding *Pseudomonas aeruginosa* polypeptide,

PT pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 31760; 455bp; English.

The invention relates to *Pseudomonas aeruginosa* polypeptides and the polynucleotides encoding them. The sequences are useful in diagnosis and therapy of pathological conditions, as molecular targets for diagnostics, prophylaxis and treatment of pathological conditions resulting from a bacterial infection, for evaluating a compound, such as a polypeptide, for the ability to bind a *P. aeruginosa* nucleic acid, as components of effective antibacterial targets, as targets for antibacterial drugs, including anti-*P. aeruginosa* drugs, as templates for recombinant production of *P. aeruginosa*-derived peptides or polypeptides, as target components for diagnosis and/or treatment of *P. aeruginosa*-caused infection, and in detection of *P. aeruginosa* sequences or other sequences of *Pseudomonas* species using biochip technology. Sequences AB067826-AB084396 represent *P. aeruginosa* polypeptides of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/Sequence.html

Query Match	10.3%	Score 147.5	DB 7	Length 1706
Best Local Similarity	24.7%	Pred. No. 0.0028		
Matches 64	Conservative 21	Mismatches 103	Indels 71	Gaps 9

Oy	19HRCRGYKVRKSYVDETLFGSPAGTRPTPPEDDPWEKEKANTRGVGEAKSLGAKG	78
Db	1277HARRASLHALART-----GLTAGRAITGEDBKRIPOARPKRTR-----	1314

Oy	79	SCEETTPASGSGTPTLTKKKNKTRPI--SHTPSYCDSELTFGSRSSGSAFCAPRMAMKGDAAK	136
Dd	1315	RANHPRGGITATPPRRPKRQDPAEGPRQAQCRQR--GSFANPASAGERR-----	136

```

Oy      137  EKAALLWPTPTPGSHSP-----RPREALRAIHAGPSKTEPCGAADSOXLMSGLHSSR  192
          |||  :      |||  :      |||  :      |||  :
Db      1365  -----PPRRRGSPARPFCCPRQRPRRGTRQSPAPTRPAED-----  1400

```

Oy 193 PLKRGLSLTLHNVPSGHPATSAHPHNGPQDLRSPSTGVTFRSPLV-----TSRARSV 247
 Db 1403 --RNNAHHPGTARPPATGRPMATAPGCAHPRHRRTATAARRCGRPSVPRATGTRRSR-- 1455

```
Qy      248 SISVSPTRRGATQKPKP 266
          : | | : | |
Db      1459 --TAPGAPQPAARQPGPRP 1475
```

RESULT 8
AAE13839

AC AAE13839;

DT 26-FEB-2002 (first entry)

DE Human lung tumour-specific protein SCC2-29

KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;

100

XX XX

2
2
0
2
3
0
0
2
1

XX

XX

PR 05-JUN-2000; 2000US-00588937.

PR 22-SEP-2000; 2000US-0234517P.

PR 01-NOV-2000; 2000US-00704512.
PR 14-DEC-2000; 2000US-00738973.
XX (CORI-) CORIXA CORP.
XX
XX
PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indrias CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
XX WPI; 2001-639201/73.
XX N-PSDB; AAD3451.
XX
XX New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e.g. lung cancer.
XX
XX Disclosure; Page 309-318; 378pp; English.
XX
XX The invention relates to isolated lung tumour-specific proteins and their
CC corresponding cDNA molecules. Lung tumour-specific proteins and their
CC antigen-presenting cells are useful for stimulating and/or expanding T
CC cells specific for a tumour protein, and for inhibiting the development
CC of cancer. The invention also relates to a composition useful for
CC stimulating an immune response, and for treating cancer. The lung tumour
CC specific oligonucleotide is useful in gene therapy and for diagnosis,
CC detection and treatment of lung cancer. The present sequence is human
CC lung tumour-specific protein
XX
SQ Sequence 4019 AA;
Query Match 9.5%; Score 136.5; DB 4; Length 4019;
Best Local Similarity 24.0%; Pred. No. 0.069;
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;
QY 29 ARTSYVDETLFGSPAGR-----PPRPDDPPVVK-----ANRTGVCK 68
DB 988 SRPLQNMETTANRPSPVRLDCSSSTTNNDPYAKPPTPRPVNDQPKSLGLSRSPVSE 1047
QY 69 EASKALGAKGSC-----ETPSRGSTPTLTPRKXK-----YRPISHTPS 108
DB 1048 QTAKGPIAAGTSDHFTKPSRADVPQQRIPDSYARLILTPALDSDGPFKTPMPPPS 1107
QY 109 YCDESLFGSRSEGASFGAPPMKADAKLRALLMTPEPTPRGSH--SPRPREAPLRAIHP 166
DB 1108 SQDP--YGSVSQ-----ASRLTSDPYERPAL--TPRPIDNFSHNSQNDPYSQPLTPHP 1158
QY 167 -----AGPSK--TEPG-----PADSQKLSMGGLHS-----SRPL 194
DB 1159 AVNESFAPHPRAFSQPETISRPTSQDPYSGPPTPRPVVDSYSSSGTARASNTDYPYQP- 1217
QY 195 KRGLSHSLTLNVPSTGHPATSAHTNGPDLRPSSTGVTFRSPPLVTSRARVSISVPSST 254
DB 1218 -----PGTPRPITVDYPSQGPQTPRPSTGTDLFTVPVTVNQHSDDPYAHHPGT 1264
QY 255 PRRGATQKPKPP 267
DB 1265 PRPGISVPSYQPP 1277
RESULT 9
ADD66733
ID ADD66733 standard; protein; 4019 AA.
XX
XX ADD66733;
XX
XX 15-JAN-2004 (first entry)
XX Human lung tumour-specific related protein, SEQ ID No 425.
XX
XX expression control; cancer; T cell; tumour; immune; cytostatic; vaccine;
KM human; lung tumour-specific.
XX Homo sapiens.
XX
XX WO200292001-A2.
5

XX
PD 21-NOV-2002.
XX
XX 10-MAY-2002; 2002MO-US014975.
XX
XX 11-MAY-2001; 2001US-00854133.
XX
XX (CORI-) CORIXA CORP.
XX
XX Lodes MJ, Wang T, Fan L, Algate PA, McNeill PD;
XX WPI; 2003-120592/11.
XX
XX New polynucleotide and polypeptide, useful for preparing a composition
PT for diagnosing, treating or preventing cancer.
XX
XX Disclosure; SEQ ID NO 425; 494pp; English.
XX
XX The invention relates to a novel isolated polynucleotide comprising one
CC of 32 47-6080 base pair sequences, given in the specification, or their
CC complements or degenerate variants, at least 20 contiguous residues of a
CC sequence in, or having at least 75 or 90 % identity with the isolated
CC polynucleotide, or that hybridise with the polynucleotide. The invention
CC further comprises: an isolated polypeptide; an expression vector
CC comprising the polynucleotide operably linked to an expression control
CC sequence; a host cell transformed or transfected with the expression
CC vector; an isolated antibody or its antigen-binding fragment that
CC specifically binds to the polypeptide; a method for detecting the
CC presence of a cancer in a patient; a fusion protein comprising the
CC polypeptide; an oligonucleotide that hybridises to the isolated
CC polynucleotide under moderately stringent conditions; a method for
CC stimulating and/or expanding T cells specific for a tumour protein; an
CC isolated T cell population; a composition comprising a first component
CC consisting of carriers and immunostimulants and a second component; a
CC method for stimulating an immune response in a patient; a method for
CC treating cancer in a patient; a method for determining cancer in a
CC patient; a diagnostic kit comprising at least one oligonucleotide or
CC antibody and a detection reagent comprising a reporter group; and a
CC method for inhibiting the development of cancer in a patient. The
CC compositions of the invention have cytostatic activity and can be used to
CC create a vaccine. The isolated polynucleotide is useful for preparing a
CC composition for diagnosing, treating or preventing cancer. This sequence
CC represents a human lung tumour-specific protein relating to the
CC invention.
XX
SQ Sequence 4019 AA;
Query Match 9.5%; Score 136.5; DB 7; Length 4019;
Best Local Similarity 24.0%; Pred. No. 0.069;
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;
QY 29 ARTSYVDETLFGSPAGR-----PPRPDDPPVVK-----ANRTGVCK 68
DB 988 SRPLQNMETTANRPSPVRLDCSSSTTNNDPYAKPPTPRPVNDQPKSLGLSRSPVSE 1047
QY 69 EASKALGAKGSC-----ETPSRGSTPTLTPRKXK-----YRPISHTPS 108
DB 1048 QTAKGPIAAGTSDHFTKPSRADVPQQRIPDSYARLILTPALDSDGPFKTPMPPPS 1107
QY 109 YCDESLFGSRSEGASFGAPPMKADAKLRALLMTPEPTPRGSH--SPRPREAPLRAIHP 166
DB 1108 SQDP--YGSVSQ-----ASRLTSDPYERPAL--TPRPIDNFSHNSQNDPYSQPLTPHP 1158
QY 167 -----AGPSK--TEPG-----PADSQKLSMGGLHS-----SRPL 194
DB 1159 AVNESFAPHPRAFSQPETISRPTSQDPYSGPPTPRPVVDSYSSSGTARASNTDYPYQP- 1217
QY 195 KRGLSHSLTLNVPSTGHPATSAHTNGPDLRPSSTGVTFRSPPLVTSRARVSISVPSST 254
DB 1218 -----PGTPRPITVDYPSQGPQTPRPSTGTDLFTVPVTVNQHSDDPYAHHPGT 1264
QY 255 PRRGATQKPKPP 267
||| : :||

Query Match 9.5%; Score 136.5; DB 7; Length 4019;

Claim 9; SEQ ID NO 1783; 1012pp + Sequence Listing; English.

CC The invention relates to an isolated polynucleotide (I) comprising a
 CC nucleotide sequence selected from any of 948 sequences (AB211119-
 CC AB212066) or their mature protein coding portion, active domain coding
 CC protein or complementary sequences. The polynucleotides are useful for
 CC identifying expressed genes or for physical mapping of human genome. The
 CC encoded polypeptides (ABP68902-ABP68949) are useful as molecular weight
 CC markers, as a food supplement, for generating antibodies, in medical
 CC imaging, screening and diagnostic assays and for treating cell-
 CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
 CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
 CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
 CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver
 CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
 CC arthritis, etc. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX

SO Sequence 4025 AA;
 Query Match 9.5%; Score 136.5; DB 5; Length 4025;
 Best Local Similarity 24.0%; Pred. No. 0.07; Mismatches 113; Indels 97; Gaps 13;
 Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARTSYDETLFGSPAGTR-----PTPPDFDPVWEK-----ANRTGCVGK 68
 Db 994 SRPLQNNMTANRSPYRDLCSSSTNNDPYAKPPDPFRPMTDQPKSLGLSRSPVSE 1053
 QY 69 EASKALAKKSC-----ETTPSRGSGTTLTPRKNK-----YRPSHTPS 108
 Db 1054 QTAGPPIAAGTSDHPTFRPADVFORQRIPDYARLILTPAPLDSPGPFKTPMQPPS 1113
 QY 109 YCDLSLGSSESGSFGAPMAKDAKRLALMTPTTPRGSH--SPRPEAPLRAIHP 166
 Db 1114 SQDP--YGSYSQ-----ASRLISVDPYRERPAL--TPPRIDNPSNOSNDPISQPLTPHP 1164
 QY 167 -----AGPSK--TEPG-----PADSQKLSMGGLHS-----SRPL 194
 Db 1165 AVNESFAPHSRASFQPGTISRPTSQDPYSQPGTPRPVVDYSQSSGTAASNTDPYSQP- 1223
 QY 195 KRGLSHSLTLNVTSGHTPATSAHTNPGQDLARSTGVTTRSLVTSRRASVSIYPSST 254
 Db 1224 -----GTPRPFTVDPYSQOPQTPRPSTQTLDFVTVTNQRHSDPYAHPGT 1270
 QY 255 PRRGATQKPKP 267
 Db 1271 PRGTSIVYSQPP 1283

RESULT 12
 ABB63299
 ID ABB63299 standard; protein; 2703 AA.
 XX
 AC ABB63299;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 16689.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 XX 27-SEP-2001.
 PD
 XX 23-MAR-2001; 2001WO-US009231.
 PF
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.

XX
 PI Venter JC, Adams M, Li FWD, Myers EW;
 XX
 DR WPI, 2001-656860/75.
 DR N-PSDB; ABLO7402.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 PS
 XX Disclosure; SEQ ID NO 16689; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
 CC sequences (ABLO1840-ABLO16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 CC
 XX

SO Sequence 2703 AA;
 Query Match 9.5%; Score 136; DB 4; Length 2703;
 Best Local Similarity 22.5%; Pred. No. 0.047;
 Matches 62; Conservative 30; Mismatches 119; Indels 64; Gaps 11;

QY 23 GGYRV-----KARTSYDETLFGSPAGTRPPPDF-----DPPVWEKANRTGCVG 67
 Db 633 GYKMGGPGQSPAGQVPQOPQOYPPGNYPRPQYRPGAYATGAPPPPTPSQA-----GAG 668
 QY 68 KEASKALGAAGSCEETTPSRGSESTTLTPRKNKXKRPISHNPSYCDLSLGSRSASGAP 127
 Db 689 GANSMBSGAQAG--GYRGRG-----MNHTGQYPPQWPPSQQIVPGAPGAMVGNH 741
 QY 128 RMAKGAAPKALRLALMTP-----PPTPRGSHSPRREAPLRAIHPAGPKTEPPAADS 180
 Db 742 VQKRG-----TPRPVVGPPPPGSSGPRPLNYLKHQIKGVCSPTPPQGP 791
 QY 181 QKLSMG--GLHSRPL--KRGLSHSLTLNVTST-----GHPATSAPIHT 220
 Db 792 QGYNGPPTGMHPMPGMPRHMGPRPHGPTMGPTSTPPOSQMLQGGQPGQASGGPES 851
 QY 221 NGPQDLRPTSGVTFRSLVTSRRASVSIYPSST 255
 Db 852 GGPBHT-SQDNGISSGGPTGAAGMHAIVSVTTGP 885

RESULT 13
 ABG70019
 ID ABG70019 standard; protein; 2703 AA.
 XX
 AC ABG70019;
 XX
 DT 05-NOV-2002 (first entry)
 XX
 DE Larval viability associated protein #18.
 XX
 KW Fruit fly; larval viability; insecticidal activity; maize; wheat; oat;
 KW rye; sorghum; rice; barley; millet; turf; cotton; sugarcane; sugar beet;
 KW oilseed rape; soybean; vegetable crop; fruit.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200257455-A2.
 XX
 PD 25-JUL-2002.
 PF 18-JAN-2002; 2002WO-US001568.
 PR 18-JAN-2001; 2001US-0262351P.
 XX

XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX Stam L, Bachmann J, Brodus J, Kamdar KP;
XX WPI: 2002-590746/63.
XX N-PSDB; ABS51395.
XX
XX Identifying inhibitors of activity of proteins essential for Drosophila
XX larval viability comprises expressing in a host a protein essential for
XX larval activity and identifying compounds that inhibit or interact with
XX the protein.
XX
XX Claim 1, Page 108-117; 169pp; English.
XX
XX The invention describes a method of identifying compounds that inhibit
XX the activity of, or that interact with a protein essential for Drosophila
XX larval viability comprising expressing in a recombinant host a DNA
XX molecule to produce a protein essential for larval viability. The method
XX is useful for identifying compounds with insecticidal activity. Compounds
XX identified are useful as insecticides in crops such as maize, wheat,
XX oats, rye, sorghum, rice, barley, millet, turf, cotton, sugarcane, sugar
XX beet, oilseed rape, soybeans, vegetable crops and fruits. This is the
XX amino acid sequence of a fruit fly larval viability associated protein
XX
XX Sequence 2703 AA;

Query Match 9.5%; Score 136; DB 5; Length 2703;
Best Local Similarity 22.5%; Pred. No. 0.047;
Matches 62; Conservative 30; Mismatches 119; Indels 64; Gaps 11;

QY 23 GGYRV-----KARSYVDETLFGSPAGNRRPPPF-----DPMVEKANRTRGVG 67
DB 633 GGYKKGSGQSGAGAGYPPQYPPGNYPPRPQYPPRAVATGPPPPPTSGA----GAG 688
QY 68 KEASKALAKGSCETTPSRGSTRPTLPRKKNYRPRISHTPSYCDSELFSGSEGAFCAP 127
DB 689 GANSWPSGAQAG--GYPRG-----MPNHTGQYPPYQWPPRPQQTVPAGAGAMVGNH 741
QY 128 RMAKDAKALRLMLTP-----PPTRGSHSPRRRPAFLAIHPAGPSKTEPPPADS 180
DB 742 VQKQ-----TPPPRVVGGPPPPGSGSPRLVYLKHQLQHKGGYGGSPTRPGP 791
QY 181 OKLSWG--GLHSRPL-----KRGLSHLTHLVNPEST-----GHPASAPHT 220
DB 792 QGYNGGPTGMFGMFWGPPHMGPPHMGPPHMGPPHMGPPHMGPPHMGPPHMGPPHMGPPH 851
QY 221 NGPQDLRPSTGVTFRSPVLTSTRASVSISVSTP 255
DB 852 GSPENI--SQDNGISSSGPTGAAGMAVTSVVTGTP 885

RESULT 14

ABO80172
ID ABO80172 standard; protein; 920 AA.
XX
XX ABO80172;
XX
XX 29-JUL-2004 (first entry)
XX
XX Pseudomonas aeruginosa polypeptide #12347.
XX
XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.
XX
XX Pseudomonas aeruginosa.
XX
XX US651795-B1.
XX
XX 22-APR-2003.
XX
XX 18-FEB-1999; 99US-00252991.
XX
XX 18-FEB-1998; 98US-0074788P.

PR 27-JUL-1998; 98US-0094190P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI: 2003-615309/58.
XX N-PSDB; ABD13743.
XX
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
XX useful as molecular targets for diagnostics, prophylaxis and treatment of
XX pathological conditions resulting from bacterial infection.
XX
XX
XX Disclosure; SEQ ID NO 28918; 455pp; English.

XX The invention relates to Pseudomonas aeruginosa polypeptides and the
XX polynucleotides encoding them. The sequences are useful in diagnosis and
XX therapy of pathological conditions, as molecular targets for diagnostics,
XX prophylaxis and treatment of pathological conditions resulting from a
XX bacterial infection, for evaluating a compound, such as a polypeptide,
XX for the ability to bind a P. aeruginosa nucleic acid, as components of
XX effective antibacterial targets, as targets for antibacterial drugs,
XX including anti-P. aeruginosa drugs, as templates for recombinant
XX production of P. aeruginosa-derived peptides or polypeptides, as target
XX components for diagnosis and/or treatment of P. aeruginosa-caused
XX infection, and in detection of P. aeruginosa sequences or other sequences
XX of Pseudomonas species using bioclip technology. Sequences ABO67826-
XX ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The
XX sequence data for this patent did not form part of the printed
XX specification but was obtained in electronic format from USPTO at
XX seqdata.uspto.gov/sequence.html

Sequence 920 AA;

Query Match 9.4%; Score 135.5; DB 7; Length 920;
Best Local Similarity 26.3%; Pred. No. 0.014;
Matches 78; Conservative 24; Mismatches 110; Indels 85; Gaps 16;

QY 5 VELAVSGNQTGLQRCGGRVAKRTSYVDETLF--GSPAGTRPTPPDPMPVVEKAN 61
DB 651 VTLRVRLQOQGVVIRDRVGLAIAAGHEY----LFAVGGQPOHQR--HPSD----- 696
QY 62 RTRGVKESKALAKGSCETTPSRGSTRPTLPRKKNYRPRISHTPSYCDSELFSGR--- 118
DB 637 ROQGPQIOAR--LGMLVSGRVRGHPAPASATRTSTGCRPPGRP--CRPSAGNRRPR 753
QY 119 -SEGASFGAPRMA-----KQDAKALRLMLTPPTPR-----GSHSPRP 156
DB 754 APQGMAGARRRRTRRGHRGSASASAGW--PPAPADRCGRAGAHSHSVGSHAPG 812
QY 157 --REAPFLAIHPAGPSKTEPPPADSQKLSMGGLHSRPLKRGLSHLTHLVNPESTGHPA 214
DB 813 AGRARPRILPMSERGPANPAPGPA-----PCARRRSR-----A 845
QY 215 TSAPTN---GPQDLRPSTGVTFRSPVLTSTRASVSISVSTPPRGATQKPPW 268
DB 846 TASCHRHAWPGMRDAGLPGAGCS--RSSYATTTDA-----PPASPRSGAMCGSRPAW 895

RESULT 15

ABP76681
ID ABP76681 standard; protein; 19938 AA.
XX
XX ABP76681;
XX
XX 26-FEB-2003 (first entry)
XX
XX Streptomyces viridochromogenes Avi gene cluster polypeptide frame 5.
XX
XX Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;
XX medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.
XX
XX Streptomyces viridochromogenes.

XX
PN W0200268436-A1.

06-SEP-2002.
PD

PF 24-AUG-2001; 2001WO-EP009815.

PR 25-FEB-2001; 2001DE-01009166.

PA (COMB-) COMBINATURE BIOPHARM AG.
PA
1978

PI Weitnauer G, Muehlenweg A, Trefzer A, Bechtold A;
...

DR WPI; 2003-018650/01.

XX
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846

PT nucleic acid encoding avilamycin synthesis enzymes.

PS Example 1; Page 68-301; 319pp; German.
 PS

CC The invention relates to avilamycin derivatives (I) with antibacterial,
CC antimycotic and fungicide activity (I) are useful for

CC virulence, protozoacidal and fungicidal (bacterial, viral, protozoal or fungal), in human
CC treatment of infections (bacterial, viral, protozoal or fungal), in human
CC or veterinary medicine, particularly where caused by *Staphylococcus*
CC aureus. (I) are more hydrophilic than known avilamyins. The present
CC sequence is that of an avilamyin synthetase enzyme from the *Streptomyces*
CC viridochromogenes Avilamyacin A biosynthetic gene cluster (AB237515-
CC AB237516)

SQ Sequence 19938 AA;

Query Match 9.4%; Score 134.5; DB 6; Length 19938;

Matches 75; Conservative 22; Mismatches 94; Indels 89; Gaps 15;

QY 49 PPDFDPWEKANRTRGVGKEASKALGAKGSC-----ETTPSR 86

Db 11439 PSELPOXWVE--RXRQSETEPMSSASSQGRCTWXRSPAQSVASKWPSANSIDSTSASR 11495

QY 87 GSTPTLTP-----RKKNKYRP---ISHTPSYCDSESLFGSRSEGASFQAPRMAGDA 134

Db 11496 SST-LVTCTSMNTNSARRRFSRPRARIS-TPG-CPDIARVSAPSSAPSSRTLAPGNA 11552

QY 135 AK-LRALW-----TPPTPRGSH--SPRPREAPLRAIHAPGSKTEPG 175

Db 11553 AKNASARSWGWLSPKSRVEAVAYRTPLYTKSFTHGSSSRTRRISRSFTVALP-----PL 11608

QY 176 PADSQKLSMGGHSSRPLKRGLSLTHLNVSPSTGHPATSAP-----HTNGPQD- 225

Db 11609 SGALPSETAWAGDHRSAPLSRATH-----CAHSGTRSPPISSKERSSAGPTEA 11659

QY 226 -----LRPSTSGVTFRSPPLVTSRARSVSI SVPSTPRGG 259

Db 11660 AHAIVRASPSASGSAQSP-KTSRASSVAVXATITSSRAG 11698

Search completed: March 28, 2005 08:51:48

Job time : 87 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2005, 08:15:06 ; Search time 27 Seconds
(without alignments)
743.726 Million cell updates/sec

Title: US-10-031-589-4

Perfect score: 1437

Sequence: 1 MKTVELAVSMQMTGLQHR.....SVSPRRRGATQKPKPKWK 269

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents, AA:*
2: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	147.5	10.3	1706	US-09-252-991A-31760	Sequence 31760, A
2	136.5	9.5	4019	US-09-854-133-425	Sequence 425, App
3	135.5	9.4	920	US-09-252-991A-28918	Sequence 28918, A
4	134.5	9.4	503	US-09-599-287A-2	Sequence 2, Appli
5	132.5	9.2	1008	US-09-252-991A-29419	Sequence 29419, A
6	131.5	9.2	713	US-09-949-016-9700	Sequence 9700, Ap
7	130.5	9.1	506	US-09-949-016-11282	Sequence 11282, A
8	130.5	9.1	705	US-09-538-092-945	Sequence 945, App
9	130.5	9.1	878	US-09-556-706B-2	Sequence 2, Appli
10	130.5	9.1	878	US-09-724-418A-2	Sequence 2, Appli
11	130	9.0	622	US-09-949-016-9104	Sequence 9104, Ap
12	128.5	8.9	501	US-09-252-991A-17730	Sequence 17730, A
13	128.5	8.9	536	US-09-252-991A-16754	Sequence 16754, A
14	128	8.9	507	US-09-599-287A-24	Sequence 24, Appl
15	126.5	8.8	693	US-09-252-991A-26071	Sequence 26071, A
16	126	8.8	696	US-08-906-865-4	Sequence 4, Appli
17	126	8.8	696	US-09-129-668-4	Sequence 4, Appli
18	125.5	8.7	699	US-10-237-551-143	Sequence 143, Appl
19	125.5	8.7	699	US-10-237-551-254	Sequence 254, App
20	125	8.7	1187	US-09-949-016-6513	Sequence 6513, Ap
21	124.5	8.7	1034	US-09-252-991A-28921	Sequence 28921, A
22	123.5	8.6	315	US-09-252-991A-16743	Sequence 16743, A
23	123.5	8.6	390	US-09-252-991A-17829	Sequence 17829, A
24	123.5	8.6	1321	US-08-317-310A-64	Sequence 64, Appl
25	122.5	8.5	1037	US-09-252-991A-17548	Sequence 17548, A
26	122	8.5	428	US-09-252-991A-25955	Sequence 25955, A
27	122	8.5	904	US-09-976-594-615	Sequence 615, App

28	121.5	8.5	309	US-09-252-991A-25386	Sequence 25386, A
29	121.5	8.5	907	US-08-783-774-2	Sequence 2, Appli
30	121.5	8.5	907	US-09-328-559A-1	Sequence 1, Appli
31	121.5	8.5	907	PCT-US95-04611A-19	Sequence 19, Appl
32	120.5	8.4	202	US-09-252-991A-32054	Sequence 32054, A
33	120.5	8.4	481	US-09-949-016-9748	Sequence 9748, Ap
34	120.5	8.4	2142	US-09-538-092-1142	Sequence 1142, Ap
35	120	8.4	345	US-09-252-991A-18076	Sequence 18076, A
36	120	8.4	405	US-09-252-991A-27573	Sequence 27573, A
37	119.5	8.3	189	US-09-252-991A-16638	Sequence 16638, A
38	118	8.2	315	US-09-252-991A-30553	Sequence 30553, A
39	118	8.2	663	US-09-252-991A-30843	Sequence 30843, A
40	117.5	8.2	432	US-09-252-991A-30848	Sequence 30848, A
41	117.5	8.2	757	US-09-949-016-7121	Sequence 7121, Ap
42	117.5	8.2	906	US-09-252-991A-32715	Sequence 32715, A
43	117	8.1	658	US-09-328-559A-2	Sequence 2, Appli
44	116.5	8.1	783	US-09-252-991A-18035	Sequence 18035, A
45	116.5	8.1	2169	US-09-949-016-6930	Sequence 6930, Ap

ALIGNMENTS

```
RESULT 1
US-09-252-991A-31760
; Sequence 31760, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31760
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31760

Query Match      10.3%; Score 147.5; DB 4; Length 1706;
Best Local Similarity 24.7%; Pred. No. 0.00015;
Matches 64; Conservative 21; Mismatches 103; Indels 71; Gaps 9;

QY      19 HRCRGYVKKARTSYVDITLFGSPAGCTPTPTDPPEVKKANRRGVGKSKALGAKG 78
DB      1277 HARRSLHALART-----GTLAGRATTEBDRRIRIQAPPRKTR----- 1314

QY      79 SCETPSGSGPTTLTPRKKNKYRPI--SHTPSYCDSELFGRSBSGASFGAPMAKDAK 136
DB      1315 RARHPGCGITPATPTPPCKQRPAPRPRQARCKQR--GSPANASAGRRR----- 1364

QY      137 LRALMTPTPPRGSHSP---RPREADLRALHPAGSEKTEPGPAADSOKLSMGSLSSR 192
DB      1365 -----PPRRGSGPAPRFRCPRRQRRPGRGPPQSPARTBPAPD----- 1402

QY      193 PLKRLSLSLTHLVNPSGHPATSPHPNGQDARPSISGTFRRPPLV-----TSRARSV 247
DB      1403 --RNAHHGTVARPPATGAPPAFTAGAHRAHRTATTAARQGPSVPRATGTRSR-- 1458

QY      248 SISVSTPRRGATQKPK 266
DB      1459 --TAPGAPQPARQPGPRP 1475

RESULT 2
US-09-854-133-425
; Sequence 425, Application US/09854133
```

Patent No. 6759508
GENERAL INFORMATION:
APPLICANT: Lodes, Michael J.
APPLICANT: Mohamath, Raedoh
APPLICANT: Henderson, Robert A.
APPLICANT: Benson, Darin R.
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.475C10
CURRENT APPLICATION NUMBER: US/09/854,133
CURRENT FILING DATE: 2001-05-11
NUMBER OF SEQ ID NOS: 735
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 425
LENGTH: 4019
TYPE: PRT
ORGANISM: Homo sapiens
US-09-854-133-425

Query Match 9.5%; Score 136.5; DB 4; Length 4019;
Best Local Similarity 24.0%; Pred. No. 0.0045;
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARTSYVETLFGSPAGTR-----PTPPDEDPWVEK-----ANRTGVCK 68
DB 988 SRPLQNMETTANRSPYRDLCSSTNNDPYAKPDTTPRVMTDQFKSLGSLSPVSE 1047
QY 69 EAKFALCAKSSC-----ETTPSGSTPLTPRKKNK-----YRDSHTPS 108
DB 1048 QTAAGPIAAGTSDHPTSPRADVFORQIPDSYARLITPAPLDSGPGPKTPMCPPTS 1107
QY 109 YCDESLSESGSFGAPYAKGDAKALMTPTPTRGSH--SPRPREAFLRAIHP 166
DB 1108 SQDP--YGSISQ-----ASRLSDPYERPAL--TPPINDFSHNSNDYISQDPLTPHP 1158
QY 167 ----AGPSK--TEPG-----PAADSQKLSMGGLHS-----SRPL 194
DB 1159 AVNESFAHPSRAFSQGTISNPTSQDPSQPGTPRVVDSYSSGSTASNDYISQF- 1217
QY 195 KRGLSHLTHLNVSTGHPSATHTNGPQDLRSTGVTFRSLVTSRARSVISVPT 254
DB 1218 -----PGRPTPTVPYSQOPQTPRPSTQTDLFVTVTQHSHPYAHPPGT 1264
QY 255 PRRGATQKPKP 267
DB 1265 PRGISVPSQPP 1277

RESULT 3
US-09-252-991A-28918
Sequence 28918, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 28918
LENGTH: 920
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28918

Query Match 9.4%; Score 135.5; DB 4; Length 920;
Best Local Similarity 26.3%; Pred. No. 0.00083;

Matches 78; Conservative 24; Mismatches 110; Indels 85; Gaps 16;
QY 5 VELAVSGMQLGIQHCRGGYRYKARTSYDELTF---GSPAGTRTPPPDPWVEKAN 61
DB 651 VVLRVRLQOQFVVHHDVRLGLAIAAGHEY---LFVVEQPOHOR--HPESD----- 696
QY 62 RTGVGKEASKALGAKGSCETTPSGSTPLTPRKKNKYRPISTHSCYDESLFSGR--- 118
DB 697 RQGGPGLQAR-LGLMVSGRVGHRAAPASATYSTGRCRPFGRF--CRPRAGNRPPR 753
QY 119 -SEGASFGAPRMA---KDAKALRALMTPTPTPR-----GSHSPRP----- 156
DB 754 APPGAMAGARRRRTRRGHRGSASASAGW-PPAPADRGCRAGCAHSPHSGVHASAG 812
QY 157 --REAPLRAIHPGPKTERGPADDSQKLSMGGLHSRLKRLSLTHLNVSTGHRA 214
DB 813 AGRAPPLPWSERGPANPAPGGA-----PCARRSR-----A 845
QY 215 TSAPHTN---GPQDLRPTSGVTFRSPPLVTSRARSVISVPTPRRGATQKPKPM 268
DB 846 TACHHAMPWGRDAGLPGAGCS-RSTATTDA-----PPASPRRSGAMCGSRPAM 895

RESULT 4
US-09-599-287A-2
Sequence 2, Application US/09599287A
Patent No. 6635446
GENERAL INFORMATION:
APPLICANT: Narayanaewamy Ramesh
APPLICANT: Ines M. Anton
APPLICANT: John H. Hartwig
APPLICANT: Ralf S. Geha
TITLE OF INVENTION: WIP, A MASP-Associated Protein
FILE REFERENCE: 1242.1022-004
CURRENT APPLICATION NUMBER: US/09/599,287A
CURRENT FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: PCT/US98/27501
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/101,457
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/068,533
PRIOR FILING DATE: 1997-12-23
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 503
TYPE: PRT
ORGANISM: Human
US-09-599-287A-2

Query Match 9.4%; Score 134.5; DB 4; Length 503;
Best Local Similarity 25.8%; Pred. No. 0.00047;
Matches 63; Conservative 23; Mismatches 107; Indels 51; Gaps 11;

QY 42 PACTR-PTPPDEDPWVEKANRTGVGKEASKALGAKGSCETTPSGSTPT--LTPRKXN 98
DB 207 PGGPRQPSGPTPPPPKNGRTALGGGSIKSPSSSFNSRPLPPTPSALDQKPP 266
QY 99 KTRPISTHSCYDESLFSGRSBGASFGAPRMAKDAKALRALMTPTPTPRGSHSPRP- 157
DB 267 PPPPVNRRPSIHREAV-----PPPPQNNKPPVPST 297
QY 158 ---EAPLRA-IHPAGSKTEPGPADSQKLSMGGLHSRLKRLSLTHLNVSTGH- 212
DB 298 PRPSAPHRPHLRPPPSRPGPPPLPS---SSGNDETPLPQKNSLSSSTPLPLSPGKS 354
QY 213 -PATSAPHTNGPQDLR--PSTSGVTFRSPPLVTSRARSVISVPTPR---RGATQK--- 263
DB 355 GRLPPPSRRPPPPVADPPGRSG-PLPPPPVSRNOSTRALPATPQLSRGVSPPRSG 413
QY 264 PKPP 267
DB 414 PRPP 417

RESULT 5
US-09-252-991A-29419

; Sequence 29419, Application US/09252991A
; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 29419

; LENGTH: 1008

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-29419

Query Match 9.2%; Score 132.5; DB 4; Length 1008;

Best Local Similarity 24.8%; Pred. No. 0.0017;

Matches 76; Conservative 25; Mismatches 141; Indels 65; Gaps 13;

QY 17 LQHRGGRY--VKARTSYDETLFGSP-AGTRPTPDPPWVEKANRTGVEKASK 72

DB 377 LVRRGRGRPRPVVAGTGHGPRRPAAGPAGADRT---DPQSGPARSARPRARRR 432

QY 73 ALGAGSCETTSRSTPLTRK--NKYRPISTPYCESLFGSSSEASGAPRM 129

DB 433 ATGAPGTQPAARQA--ALRPPGAGAGAFLLGCRPTAAQAPGRTGAPRSRRV 490

QY 130 AKGDAXLRALLMTPTPT--PRGSHPRPREAPLRAIHAGSKTEPGPAASOKLSMG 187

DB 491 AAGLRKRRQPARLRPPPTQFPGSSGDPARAPRRRLALPARPELDADPARTRQRRACR 550

QY 188 LHS-----RPLKGLSHSLTLNVPSTGH-----PATSAPTN 221

DB 551 RHGAGODLAVAGPRAPKAG--RAARHAGAGDAHQDPQLAGRALRPARABAPAR 608

QY 222 GPQD-LRP--STSGVTRSPVTSRARSVSISV-PSIPRRG-----G 259

DB 609 RPAPRLRPHRRARPGADHLCPAATRRRRRAGAAVPPADPRGAHQERHHQGRRRPRAG 668

QY 260 ATQKPKP 266

DB 669 GTASFPV 675

RESULT 6

US-09-949-016-9700

; Sequence 9700, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: C1001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; PRIOR FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 9700

; LENGTH: 713

; TYPE: PRT

; ORGANISM: Human

US-09-949-016-9700

Query Match 9.2%; Score 131.5; DB 4; Length 713;

Best Local Similarity 25.2%; Pred. No. 0.0014;

Matches 61; Conservative 18; Mismatches 110; Indels 53; Gaps 8;

QY 40 GSPACTRPTPDPPWVEKANRTG-----VKEASALGAKSCETTSRGSTP 90

DB 462 GPAPQGR-PPQGGPQPGPGRGPPPLQGRPPPPQGGHSLGCLPPAG-SPLPRLSP 519

QY 91 TLTPKKNKVRPISTPYCESLFGSSSEASGAPRMAGDAKALALMTPTPTPR- 149

DB 520 TSAFQ-----QASQAPPTQGGKRSKPVVAGGPAAPARPPSPQORAGPQATRQ 574

QY 150 ---GSHSPRPREAPLRAIHAGSKTEPGPAASOKLSMGSLHSSRLKGLSHSLTHL 205

DB 575 TVSGPAPRKASGAPPGGQGRGPPQKPPGAPGTRQASQAG----- 616

QY 206 NVPTGHPATSAPTNGPQDLRPTSGVTRSPVTSRARSVSISVSTPRRGATQKPK 265

DB 617 PVPRTPPTTQPP-----RPSGPGRGRPRKPOLAQKPSQDVPPATAAAG----- 662

QY 266 PP 267

DB 663 PP 664

RESULT 7

US-09-949-016-11282

; Sequence 11282, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: C1001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; PRIOR FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11282

; LENGTH: 506

; TYPE: PRT

; ORGANISM: Human

US-09-949-016-11282

Query Match 9.1%; Score 130.5; DB 4; Length 506;

Best Local Similarity 25.0%; Pred. No. 0.0011;

Matches 61; Conservative 24; Mismatches 108; Indels 51; Gaps 10;

QY 42 PAGTR-PTPPDPPWVEKANRTGVEKASKALGAKSCETTSRGSTPT--LTPRKXN 98

DB 210 PGGPRQPGPTPTPPFPANRGITAGGSSIRSPSSSPSRRNRPPLPTPSBALDDKPP 269

QY 99 KYRPISTPYCESLFGSSSEASGAPRMAGDAKALALMTPTPTPGS-----HS 153

DB 270 PPPPVGNRPSIHREAV-----PPPPQNNKPPVPST 300

QY 154 RPPEAPLRAIHAGSKTEPGPAASOKLSMGSLHSSRLKGLSHSLTHLNVSTGH- 212

DB 301 PRPSASQAPPPPPSPRPPPLPPS---SSGNDPTFRLPQRNLSSSTPPPLSPGSR 357

QY 213 -PATSAPTNGPQDLR--PTSGVTRSPVTSRARSVSISVSTPR---RGATQK--- 263

Db 638 VTSOPKNAATVTTGQHNRPSNDELTSTSDNSTSHMGENTITQVTPASISTHNVST 697
Qy 144 PEPTP-----GSHSPRPREA-----PLRAIHAGS-----KTEBP 176
Db 698 SPSPEPCTTSQASGPPNSSTKPGENVVTKTPPQNTSQASGQKTAVPVYTSSTG 757
Qy 177 AADSKLMSGLHSSRPPLKRGSLHSLTHLVNST---GHPATSAPHTNGPQDLRPSISGV 233
Db 758 KANS---TTGGKHTT-----GHGARTSEPTTDYGGDSTTPRRRYNATYTLRPSSTSK 807
Qy 234 -----TFRSPLVTSRARSYSISVPSTPR 256
Db 808 LRPRWTFTSPVTTAQTAVPVPPTSQPR 835

RESULT 11
US-09-949-016-9104
; Sequence 9104, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: fastseq for windows version 4.0
; SEQ ID NO 9104
; LENGTH: 622
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9104

Query Match 9.0%; Score 130; DB 4; Length 622;
Best Local Similarity 29.6%; Pred. No. 0.0016;
Matches 74; Conservative 27; Mismatches 81; Indels 68; Gaps 18;

Qy 67 GKESKALGAK-GSCETTPSRGSTPTLTPRKKNKTRIPSHPSYCDSELSFGSRSGASFG 125
Db 19 GPLAQTGGARIGS---SPSFTSTMDLTP-PRKYNPLR-----NESI-SLLEFGASGS 67
Qy 126 AP--RMAGDPAKRLALMTPTPTRGSHSPR-----PREAPLRAIHAG--PSKTEP 174
Db 68 TPPEELPSPASSLPIL--PLLP-GDDSPPTLCSFPFRKSNLRLANPAGRPSGKEP 123
Qy 175 GPADS-----OKLSMGLHSSRPPLKRG-----LSHSLTHLN 206
Db 124 GRAADDGSGIYGAAMPDSGLPLLDPMKLSGGGRRTTR-VEGGLGGEEMTRHG-SFVN 181
Qy 207 VPSTG--HP--ATSAAPHNN-----GPODLRPSISGVTFRSPVTSRARSYSISVPSTPR 256
Db 182 KPTRWMLPNDKVMGPVSYLVRYMGCVEVLQSKBALDFNTQTQVTR-EAISLVEAVPG 240
Qy 257 RGATQXPKP 266
Db 241 AKGATRRKKP 250

RESULT 12
US-09-252-991A-17730
; Sequence 17730, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17730
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17730

Query Match 8.9%; Score 128.5; DB 4; Length 501;
Best Local Similarity 24.3%; Pred. No. 0.0016;
Matches 67; Conservative 25; Mismatches 113; Indels 71; Gaps 12;

Qy 2 KTVPELAVSGMQTGLQHRGGRGVKARTSYVDETLFGSPAG-----TRPTPDF 52
Db 274 RTPV-----RGRGRTARPPAP-----LAAAGPALPARRQRGRGPGI 314
Qy 53 DPWVEKANRTRGVGEKASKALGAKSGCTTPSRGSTPTLTTPRKKNKTRIPSHPSYCD 112
Db 315 -----PQNPQRPVK---RATGQRQVRRPPARCAGAVR-RGRGDRPAPR----- 357
Qy 113 SLFGSRSGASFGAPRAKDAKRLALMTPTPTRGSHSPRPREAPLRAIHGPKST 172
Db 358 ---GTRGGQAAVRRHRGTGA-----PDRACGQOBERPRO---RAGQPAHPARG 402
Qy 173 EPPGADSGKLSMGLHSSRPPL-----KR---GLSHSLTHLVNSTGHATSAPHING 222
Db 403 RPPARQGLPAPRGDPHGRRLTPGNGCAKEAMCIPHGFHF-ISSGHNGEWPGRDGS 461
Qy 223 PDLRPSISGVTFRSPVTSRARSYSISVPSTPRG 258
Db 462 PRSPAPAPRRARRRGATIALRLIDLNRAGREG 497

RESULT 13
US-09-252-991A-16754
; Sequence 16754, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16754
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16754

Query Match 8.9%; Score 128.5; DB 4; Length 536;
Best Local Similarity 21.9%; Pred. No. 0.0018;
Matches 87; Conservative 27; Mismatches 113; Indels 171; Gaps 19;

Qy 18 QHRC---RGRVYKARTSYVDETLFGSPAGTRP-----TPPDPPPWVEKANRTRG 65
Db 81 RHRCLPASATYRSTCKTISAASP---SCAGSTPRXMRRAAATPSSCAGTDQSRKTS 137
Qy 66 -----VGKESKALGAKSGCTTPSRGSTPTLTTPRKKNK-----YRPI-----SHTPSYC 110
Db 138 ASGCVSNNAAR---QADSCAASGRRARRKSLQCRNRNRIGTGRPLMAASQASQLHRYH 194

```

QY      111 DESLFGSR-----SEGSF 124
Db      195 GGNRRARRRRSCKRALSTSSMARLAARSSSSMPPPMRRNRATRTTRAGSGCGST 254
QY      125 GAPPMAGDKDAKTLALMTPTPTPGSGHSPRPREAPRLAHPA-----167
Db      255 GGGVTGTSSASPRPA--TPAAGAGARRARPRRPARTRCPACRCGSAAPRRHRRR 312
QY      168 -----GPEKTEPG-----PAADSQLSMGGLHSSRPLKGL 198
Db      313 DRRRPSGGCGGPGTSPGVRVGRVAGNCARAGPAGAARRARDG-----RPVRR-- 365
QY      199 SHSLTHLVNSTGTPATAP--HTNGPQDLRPTSGVTFRSPVLTSPAR---SVSISVPS 253
Db      366 -RALAGCPVPAAGRPGRSRRPARNSGSP--RPPVAG--RDPSPGRALAAAGSVADSPG 418
QY      254 T-----PRRG-----ATOKPRP 266
Db      419 TRRARRAAGCGRRRRARKRAVPRRSGPCRPAATGAPPP 456

```

RESULT 14

```

US-09-599-287A-24
; Sequence 24, Application US/09599287A
; Patent No. 6635446

```

```

; GENERAL INFORMATION:
; APPLICANT: Narayanasamy Ramesh
; APPLICANT: Ines M. Anton
; APPLICANT: John H. Hartwig
; APPLICANT: Ralf S. Geha
; TITLE OF INVENTION: WIP, A WASP-Associated Protein
; FILE REFERENCE: 1242.1023-004
; CURRENT APPLICATION NUMBER: US/09/599, 287A
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: PCT/US98/27501
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/101,457
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/066,533
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translated WIP ORF No. 6635446 3
US-09-599-287A-24

```

```

Query Match      8.9%; Score 128; DB 4; Length 507;
Best Local Similarity 25.3%; Pred. No. 0.0018;
Matches 62; Conservative 23; Mismatches 108; Indels 52; Gaps 11;

QY      42 PAGTR-PTRPPDPPWVEKANKRTGCGKASKALGAKSSCETTSRSGSTP--LTPRKYN 98
Db      210 PGGRRQPGSGTPPPFPNNGTALGGGSIROSPLSSSPFNRRPLPTPSRALDDKPP 269
QY      99 KYRPISTPTSYCDSESLFGSRSEGSFGAPMAKGDAAKALALLTPTPTPGSGHSPRR- 157
Db      270 PPPVGNRRPSIHRNAV-----PPPPQNNKPPVPVST 300
QY      158 ---EAPLRA-IHPAGPSTKTEPGPADSOKLMSGILHSSRPLKGLSHSLTHLVNSTGHP 213
Db      301 PRPSAPHRPRLRPPSPRRGPPPLPPS---SSGNDETRLPQRLNLSLSSSTPLPSPGRS 357
QY      214 A---TSADHTNGPQDLR--PSTSGVTFRSPVLTSPARSVSISVSTPR---RGATOK-- 263
Db      358 GPLPPPVSESRPPVRRPGRSG-PLPPPPVSRNGSTRALPATPQLPSRSGVDSPRS 416
QY      264 -PKPP 267
Db      417 GPRPP 421

```

RESULT 15

```

US-09-252-991A-26071
; Sequence 26071, Application US/09252991A
; Patent No. 6551795

```

```

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfeld et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252, 991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26071
; LENGTH: 693
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26071

```

```

Query Match      8.8%; Score 126.5; DB 4; Length 693;
Best Local Similarity 26.9%; Pred. No. 0.0038;
Matches 87; Conservative 28; Mismatches 112; Indels 97; Gaps 19;

```

```

QY      16 GLQHRCC-----GGYRKATSYVDLTFGSP---AGTRPTPPDPPWVEKANKRR 64
Db      332 GLPQRAKOPAAATGAGPTAGRQRSSGADP--GQPLQAGGRLEP-----NVRP 378
QY      65 GVGKESKALGAKSGCETTPSRGSTPTLP---RKNKYRPISHTPSYCDSESLFGSRSE 120
Db      379 GVALRRRRARTGSGVGNPPAR--YDILRPRGARGQLHPRPGAGDLS-CDRPAFRRAR 435
QY      121 GASFGAP---RMAKD---AAKLRLMT-----PPTRGSH 152
Db      436 TALGAPAAATQVAGRRDGRSALLRALQADPTHGRRRRGRRRQVPRQSPAPPRGDR 495
QY      153 S-PPREPAP-----LRLIHPAGPSTKTEPGA---ADQKLSMGGLHSSRPL 194
Db      496 TGRPADPAATGALPGRPSGRLRPH--AGPG--HPGARGRPGRADHPQPLAGLDPRPR 552
QY      195 KR-GLSHSLTHLVNSTGHPATGAPTNGP-----ODLRPSTSGVTFRSPVLTYSRA 244
Db      553 SRPGVQKGLCRAGIPGTAMPAT---HAGPAAASLBERGAAYLPAGLGLRRRTRRNAARS 609
QY      245 RSVSISVPS---TPRRGATOKP 264
Db      610 RRGAGQGPSRGLRRRRGRARRLP 633

```

```

Search completed: March 28, 2005, 08:54:14
Job time : 30 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2005, 08:53:20 ; Search time 390 Seconds
(without alignments)
228.375 Million cell updates/sec

Title: US-10-031-589-4

Perfect score: 1437

Sequence: 1 MKTPVELAVSGMQLGLQHR.....SVSPRRRGATQKPKPKWK 269

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1407402 seqs, 331100923 residues

Total number of hits satisfying chosen parameters: 1407402

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database :

Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/2/pubpaa/PCT_NRW_PUB.pep:*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*

7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*

10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*

11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*

12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*

13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*

14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*

15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*

16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*

17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*

18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*

19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1216	84.6	229	US-10-452-858C-79	Sequence 79, Appl
2	239	16.6	467	US-10-276-774-1711	Sequence 1111, Ap
3	218	15.2	65	US-09-864-408A-8426	Sequence 8426, Ap
4	150	10.4	668	US-10-437-963-134726	Sequence 14726,
5	139	9.7	429	US-10-437-963-114012	Sequence 14012,
6	136.5	9.5	4019	US-09-738-973-425	Sequence 425, App
7	136.5	9.5	4019	US-09-854-133-425	Sequence 425, App
8	136.5	9.5	4019	US-10-144-649A-425	Sequence 425, App
9	135	9.4	442	US-10-425-114-40452	Sequence 40452, A
10	134.5	9.4	503	US-10-078-547-2	Sequence 2, Appli
11	134.5	9.4	19652	US-10-084-846A-7	Sequence 7, Appli
12	134	9.3	625	US-10-437-963-143626	Sequence 143626,
13	134	9.3	937	US-10-437-963-156445	Sequence 156445,

14	133	9.3	380	US-10-437-963-200034	Sequence 200034,
15	133	9.3	514	US-10-437-963-111701	Sequence 111701,
16	132.5	9.2	488	US-10-365-742-20	Sequence 20, Appl
17	132.5	9.2	497	US-10-437-963-125004	Sequence 125004,
18	132	9.2	621	US-10-437-963-177697	Sequence 177697,
19	131.5	9.2	277	US-10-437-963-177998	Sequence 177998,
20	130.5	9.1	417	US-10-437-963-143835	Sequence 143835,
21	130.5	9.1	878	US-10-722-050-2	Sequence 2, Appli
22	130	9.0	307	US-10-437-963-181279	Sequence 181279,
23	130	9.0	391	US-10-437-963-187780	Sequence 187780,
24	129.5	9.0	791	US-10-170-385-57	Sequence 57, Appl
25	129.5	9.0	863	US-10-359-012-2	Sequence 2, Appli
26	129.5	9.0	863	US-10-359-012-14	Sequence 14, Appl
27	129.5	9.0	19723	US-10-084-846A-5	Sequence 5, Appli
28	128	8.9	507	US-10-078-547-24	Sequence 24, Appl
29	128	8.9	514	US-10-437-963-195781	Sequence 195781,
30	128	8.9	797	US-10-156-761-10907	Sequence 10907, A
31	127.5	8.9	388	US-10-437-963-160033	Sequence 160033,
32	127.5	8.9	713	US-10-437-963-150342	Sequence 150342,
33	127.5	8.9	1870	US-10-408-765A-120	Sequence 120, App
34	127	8.8	238	US-10-437-963-145197	Sequence 145197,
35	127	8.8	481	US-10-437-963-176449	Sequence 176449,
36	126.5	8.8	1321	US-10-694-874-4	Sequence 4, Appli
37	126	8.8	696	US-09-129-668-4	Sequence 4, Appli
38	126	8.8	696	US-10-122-805-4	Sequence 4, Appli
39	125.5	8.7	699	US-10-121-888-143	Sequence 143, App
40	125.5	8.7	699	US-10-200-562-143	Sequence 143, App
41	125.5	8.7	699	US-10-237-551-143	Sequence 143, App
42	125.5	8.7	699	US-10-237-551-254	Sequence 254, App
43	125.5	8.7	1103	US-10-437-963-166325	Sequence 166325,
44	125	8.7	231	US-10-767-701-18778	Sequence 18778, A
45	125	8.7	469	US-10-437-963-136216	Sequence 136216,

ALIGNMENTS

RESULT 1

US-10-452-858C-79

Sequence 79, Application US/10452858C

Publication No. US20040086945A1

GENERAL INFORMATION:

APPLICANT: Sreekrishna, Kotikanyadnam

APPLICANT: Goetner, Daniel R.

TITLE OF INVENTION: HAIRLESS PROTEIN-INTERACTING PARTNER COMPLEXES AND METHODS THEREFOR

FILE REFERENCE: 8956P

CURRENT APPLICATION NUMBER: US/10/452,858C

NUMBER OF SEQ ID NOS: 94

SOFTWARE: PatentIn version 3.1

SEQ ID NO 79

LENGTH: 229

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

LOCATION: (217)..(217)

OTHER INFORMATION: The 'Xaa' at location 217 strands for Lys, Arg, Thr, or Met.

FEATURE:

NAME/KEY: misc feature

LOCATION: (219)..(219)

OTHER INFORMATION: The 'Xaa' at location 219 strands for Tyr, Cys, Ser, or Phe.

FEATURE:

NAME/KEY: misc feature

LOCATION: (225)..(225)

OTHER INFORMATION: The 'Xaa' at location 225 strands for Leu, or Phe.

US-10-452-858C-79

Query Match 84.6%, Score 1216, DB 15, Length 229,

Best Local Similarity 98.7%, Pred. No. 8.2e-83,

Matches 226, Conservative 0, Mismatches 3, Indels 0, Gaps 0,

Qy	1	GMOITLGIQHRCRGVYVKARTSYVBETLLFGSAGNRPRPPDPDPMWZKARTEGVCKEA	70
Dz	1	GMOTLGLIQHCRGGRGYVKARTSYVBETLLFGSAGNRPRPPDPDPMWZKARTEGVCKEA	60
Qy	71	SKALGAKGSCETPPSRGSTPTLLPRKKNCYRPISHTPSYCDSLLFGSRSEGAFCABPMA	130
Dz	61	SKALGAKGSCETPPSRGSTPTLLPRKKNCYRISHTPSYCDSLLFGSRSEGAFCABPMA	120
Qy	131	KGDAAKIRALLMTPPTPPRGSHSPREAPLPAIHPIAGBSKTPEGDAOSOKLSMGILHS	190
Dz	121	KGDAAKIRALLMTPPTPPRGSHSPREAPLPAIHPIAGBSKTEPEPADSOIKLSMGILHS	180
Qy	191	SRLPLXKGLSHSLTHLVNPSGTGHAPATSGAHPNTNGPOLRSTSGVYFRSPYL	239
Dz	181	SRLPLXKGLSHSLTHLVNPSGTGHAPATSGAHPNTNGPOLRXYSGVYXRSPYL	229

RESULT 2

```

US-10-276-774-1711
: Sequence 1711, Application US/10276774
: Publication No. US20040053245A1
: GENERAL INFORMATION:
: APPLICANT: Hyseq, Inc.
: APPLICANT: Tang, Y, Tom et al
: TITLE OF INVENTION: No. US20040053245A1 Nucleic Acids and Polypeptides
: FILE REFERENCE: 21272-030
: CURRENT APPLICATION NUMBER: US/10/276,774
: PRIOR FILING DATE: 2002-11-18
: PRIOR APPLICATION NUMBER: 09/560,875
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: 09/496,914
: PRIOR FILING DATE: 2000-02-03
: NUMBER OF SEQ ID NOS: 2700
: SOFTWARE: Custom
: SEQ ID NO 1711
: LENGTH: 467
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURES:
: NAME/KEY: misc_feature
: LOCATION: (1)...(467)
: OTHER INFORMATION: Xaa = any amino acid or nothing
: US-10-276-774-1711

```

Query Match 16.68; Score 239; DB 15; Length 467;

Best Local Similarity 29.4%; Pred. No. 1.1e-09;
Matches 94; Conservative 30; Mismatches 94; Indels 102; Gaps 17;

[illegible]

RESULT 3

```

US-09-864-408A-8426
Sequence 8426, Application US/09864408A
Publication No. US20040009474A1
GENERAL INFORMATION:
APPLICANT: Leach, Martin D.
APPLICANT: Shinkets, Richard A.
TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encoded by
FILE REFERENCE: 21402-012
CURRENT APPLICATION NUMBER: US/09/864,408A
CURRENT FILING DATE: 2001-05-24
PRIOR APPLICATION NUMBER: 60/206,690
PRIOR FILING DATE: 2000-05-24
NUMBER OF SEQ ID NOS: 9068
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8426
LENGTH: 65
TYPE: PRT
ORGANISM: Homo sapiens
US-09-864-408A-8426

```

Query Match	15.2%;	Score 218;	DB 11;	Length 65;
-------------	--------	------------	--------	------------

Best Local Similarity 71.4%; Pred. No. 4.4e-09;
Matches 40; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

```

Qy      12  MÖTLGLOHRCRGYRVKARTSYVDELFGSPAGTRPTPPDFDPWEKANTRGVG  67
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1  MKAITHLOHRSPTSYRVKAPASYVDELFGSPARTRPQPDFDPWVQNCNRSRGVG  56

```

RESULT 4

```

US-10-437-963-134726
/ Sequence 134726, Application US/10437963
/ Publication No. US20040123343A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovallig, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Wu, Wei
/ APPLICANT: Boukharov, Andrey A.
/ APPLICANT: Barbazuk, Brad
/ APPLICANT: Li, Ping
/ TITLE OF INVENTION: Rice
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53221) B
/ CURRENT APPLICATION NUMBER: US/10/437,963
/ CURRENT FILING DATE: 2003-05-14
/ NUMBER OF SEQ ID NOS: 204966
/ SEQ ID NO 134726
/ LENGTH: 668
/ TYPE: PRT
/ ORGANISM: Oryza sativa
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT4530_36471C.1.dep
/ US-10-437-963-134726

```

Query Match 10.4%; Score 150; DB 16; Length 668;

Best Local Similarity 26.7%; Pred. No. 0.0068;
Matches 72; Conservative 21; Mismatches 113; Indels 64; Gaps 13;

OY 23 GGVKARTSYDETLFGSPAGTRPPEPDFD - PWVEKARTRGVGEKASKALGAKSCB 81
 ||| :
Dy 80 GGYNPPSPICISPTTPGGGGGYPTPBDFTPSSSSTPSSTPGCGSSPT -----PCD 134
 ||| :
OY 82 TTPSRGS-TPLTTPKKNNRYPRISHTPSYCDESIFGRSGEASFGAPRAKGDAAKLRL 140
Dy 135 APPSSSDTSPTTPGGGGYSPT--TFs-----DTPPSSD----- 168
 ||| :
OY 141 LMTPEPTP-RGSHSPPREAP-----LRAIHPAG-----PSKTEPPAADSOGLSM 185
 ||| :
Dy 169 --TSPTTPGGGGYPTPDSADAPPSSSDTSPTTPGGGGGYTPTPSDADAPPSSSDTSPTTP 226
 ||| :

QY 186 GGLHSSRLPKRGLSHSLTHLNVSTGHPATNSAPHTNGPDLRSTST-----GVTFR 236
DB 227 GGGGGYTPPSPDTPPSPSSGSSPTT--PGGGGGYTPPSPDTPPSPSSGSSRTTPGCGSTP 284
QY 237 SPLVTSRARSVSISVSPSTPRRGATOKKPP 266
DB 285 TPCGTPPAPSSGTS-PTTP--GGSYYPPTP 311

RESULT 5

US-10-437-963-114012
; Sequence 114012, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 114012
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(429)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_17745C.1.pcp
US-10-437-963-114012

Query Match 9.7%; Score 139; DB 16; Length 429;
Best Local Similarity 24.5%; Pred. No. 0.027;
Matches 83; Conservative 36; Mismatches 100; Indels 120; Gaps 19;

QY 21 CRGGR---YKARSTYDETLFGS-----PAG-----TRPPPPD-----FPPVVEKAN 61
DB 5 CTGVPRWPAFYRRRRTYRQLLTORSPHPPPAALVVSQPPPDAPLAPIRPISIGAS 64
QY 62 RTRGVGKEASKALGAKGS-----CETTPSRG-----STPPLTPRKKNKY 100
DB 65 RAKPASISAAARSRAEPAPAICTSTPSRSPIPPNDAMPARCNRLPPRPLAVSRHAS 124
QY 101 RPISH-----TPSYCDESLFGSRSEGASFGAPMAKGDAAKLRLALMTPTPPRSGH 152
DB 125 NDRSHHNSRCCQMSHLAIALITS--SAAATKRSRPRSP--QAVRCRSL---QPPAP----- 175
QY 153 SPRPREADPLRAIHPPAGPSKTEPFPAPADSOKLSMGSLH-----SSRPLKKGSLH- 200
DB 176 -----AVAARELHVAGRDSTLPP-----HVAGLHDACITAGLAHEAARPPRCLDHH 223
QY 201 -----SLHLNVSTGHPATSA--PHTNGPDLRSTSTGYTPRSL--VTSRAR----- 245
DB 224 GRCLASQSPQSMPPQPPVTAAGVSHNRQPL-----TGAVVASSCIAGVTGSRXKPPPP 280
QY 246 -----SVSISVSPSTPRRGATOKKPP 267
DB 281 HLSGSRASVLTGGSSGSVSKVATAFGR-----RRPKPP 315

RESULT 6
US-09-738-973-425
; Sequence 425, Application US/09738973
; Patent No. US20020110563A1

; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 425
; LENGTH: 4019
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-973-425

Query Match 9.5%; Score 136.5; DB 9; Length 4019;
Best Local Similarity 24.0%; Pred. No. 0.5;
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARTSYVDETLFGSPVAGTR-----PTPPDPFPVWEK-----ANRRGVGX 68
DB 988 SRPLQWNETTANRSPVAVDLCCSSTTNDPYAKPPDTPRPVWTDQFPKSLGSRSPVSE 1047
QY 69 EASKALGAKGSC-----ETTPSRGSTPPLTPRKKNK-----YRPISHTPS 108
DB 1048 QTAKEPIAAGTSDHTKSPRADVGRQRIIDSVARPLTLTAPLDSGCGPFTPMQPPS 1107
QY 109 YCDBSLFGSRSEBGSFGAPRAKGDAAKLRLALMTPTPPRSGH--SPRPREADPLRAIHP 166
DB 1108 SQDP--YGSVQ-----ASRLISVDPYERPAL--TPRIDNFSHNSQNDPYSQPLTPEHP 1158
QY 167 -----AGPSK--TEPG-----PADSOKLSMGSLH-----SRPL 194
DB 1159 AVNESFAHSPASFGISPTSDPYSPQPPTRPVVDYSQSGTARSSTDPYSOP- 1217
QY 195 KRGSLSHLTHNVSTGHPATNSAPHTNGPDLRSTSTGYTPRSL--VTSRAR----- 254
DB 1218 -----POTRPPPTVDVYSQPPQTPRSTOTDLFTVPTVNTGNRHDVYHPPTGT 1264
QY 255 PRRGATOKKPP 267
DB 1265 PRPGISVPSQPP 1277

RESULT 7

US-09-854-133-425
; Sequence 425, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 425

LENGTH: 4019
TYPE: PRT
ORGANISM: Homo sapiens
US-09-854-133-425

Query Match 9.5%; Score 136.5; DB 9; Length 4019;
Best Local Similarity 24.0%; Pred. No. 0.5;
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARTSYVDETLFGSPAGTR-----PTPPDDPPWVEK-----AARTRGVCK 68
DB 988 SRPLQNMETIANRSPVRLDCSSSTNNDPYAKPDTTPRVMTDQPKSLGLSRSPVSE 1047
QY 69 EASKALGAKGSC-----ETTPSRGSTRPTLTPRKKNK-----YRPSHTPS 108
DB 1048 QTAKGPIAAGTSDHFTKPSRADVFCQRIPDSYARLITPAPLDSPGPFKTPMQPPS 1107
QY 109 YCESLFGSRSEGSFAPPMKADAKLRALLMTPTPRGSH--SPRREAPLRAIHP 166
DB 1108 SQDP--YGSVSQ-----ASRRLSVDPYERPAL--TPRPIINFSHNSQNDPYSQPPLTPHP 1158
QY 167 -----AGPSK--TEPG-----PADSQKLSMGGLHS-----SRPL 194
DB 1159 AVNESFAHPSRARSQGTISRPTSQDPYQPGTTPRVVDSYSSQSGTASNTDPYSQP- 1217
QY 195 KRGLSHLTHLVNSTGHPATSAPTNGPDLRPTSTGVTFRSPPLVTSRARSVISVPS 254
DB 1218 -----FGTPRPTTVDPYSQQTTPRSTQTLFVTPVTVNQRHSDPYAHHPGT 1264
QY 255 PRRGATQKPKPP 267
DB 1265 PRGIVSVYSQP 1277

RESULT 8
US-10-144-649A-425
Sequence 425, Application US/10144649A
Publication No. US20030118599A1
GENERAL INFORMATION:
APPLICANT: Lodes, Michael J.
APPLICANT: Wang, Tongtong
APPLICANT: Fan, Liqun
APPLICANT: Algate, Patricia A.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 210121.475C11
CURRENT APPLICATION NUMBER: US/10/144,649A
CURRENT FILING DATE: 2002-08-21
NUMBER OF SEQ ID NOS: 749
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 425
LENGTH: 4019
TYPE: PRT
ORGANISM: Homo sapiens
US-10-144-649A-425

Query Match 9.5%; Score 136.5; DB 14; Length 4019;
Best Local Similarity 24.0%; Pred. No. 0.5;
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARTSYVDETLFGSPAGTR-----PTPPDDPPWVEK-----AARTRGVCK 68
DB 988 SRPLQNMETIANRSPVRLDCSSSTNNDPYAKPDTTPRVMTDQPKSLGLSRSPVSE 1047
QY 69 EASKALGAKGSC-----ETTPSRGSTRPTLTPRKKNK-----YRPSHTPS 108
DB 1048 QTAKGPIAAGTSDHFTKPSRADVFCQRIPDSYARLITPAPLDSPGPFKTPMQPPS 1107
QY 109 YCESLFGSRSEGSFAPPMKADAKLRALLMTPTPRGSH--SPRREAPLRAIHP 166
DB 1108 SQDP--YGSVSQ-----ASRRLSVDPYERPAL--TPRPIINFSHNSQNDPYSQPPLTPHP 1158

QY 167 -----AGPSK--TEPG-----PADSQKLSMGGLHS-----SRPL 194
DB 1159 AVNESFAHPSRARSQGTISRPTSQDPYQPGTTPRVVDSYSSQSGTASNTDPYSQP- 1217
QY 195 KRGLSHLTHLVNSTGHPATSAPTNGPDLRPTSTGVTFRSPPLVTSRARSVISVPS 254
DB 1218 -----FGTPRPTTVDPYSQQTTPRSTQTLFVTPVTVNQRHSDPYAHHPGT 1264
QY 255 PRRGATQKPKPP 267
DB 1265 PRGIVSVYSQP 1277

RESULT 9
US-10-425-114-40452
Sequence 40452, Application US/10425114
Publication No. US20040034888A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovacic, David K.
APPLICANT: Tabacka, Steven E.
APPLICANT: Tabacka, Jack E.
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(5313) B
CURRENT APPLICATION NUMBER: US/10/425, 114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 40452
LENGTH: 442
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: LIB143-063-EB_FLI.pep
US-10-425-114-40452

Query Match 9.4%; Score 135; DB 15; Length 442;
Best Local Similarity 28.5%; Pred. No. 0.056;
Matches 69; Conservative 28; Mismatches 101; Indels 44; Gaps 12;

QY 37 TLFGSPAGTRPTPPDDPPWVEKANRTRGVGEASKALGAKGSCETTPSRGSTRPTLTPR- 95
DB 55 TTRRCRGISPSPTSPSSAPSR-----PRPATSPPTGACRCPHP- 102
QY 96 KKKKYPISHTPSYCESLFGSRG--EGASFGAPPMKADAKLRALLMTPTPRGSHSP 154
DB 103 RRSRRRPPSSAS-----ASGAAATAAFA-APSATATSPCTSPR--SP 145
QY 155 RPREAPLRAIHPGPKTEPG--PADSQKLSMGGLHSRPLKRLSLTHLVNSTG 211
DB 146 -TRTPTRASRSPARSFGSPSPPTPTASGTTTSRSIAR--CSTBACGRCPTSG 203
QY 212 HPATSAH--TNGPDLRPT--SGVTFRSP-----LVTSRARSVISVSTPRRGATOK 263
DB 204 ACATWSPSTCTPARATPTGTRGSSAPSPASSSTASTSSRVGATCATSASRSRSTR 263
QY 264 PK 265
DB 264 PR 265

RESULT 10
US-10-078-547-2
Sequence 2, Application US/10078547
Publication No. US20020199211A1
GENERAL INFORMATION:
APPLICANT: Narayanasamy Rameeh
APPLICANT: Ines M. Anton
APPLICANT: Miguel A. de la Fuente
APPLICANT: Ralf S. Geha
TITLE OF INVENTION: WIP, A WASP-Associated Protein

FILE REFERENCE: 1242.1022-005
CURRENT APPLICATION NUMBER: US/10/078,547
CURRENT FILING DATE: 2002-02-19
PRIOR APPLICATION NUMBER: 09/599,287
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: PCT/US98/27501
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/101,457
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/068,533
PRIOR FILING DATE: 1997-12-23
NUMBER OF SEQ ID NOS: 27
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 2
LENGTH: 503
TYPE: PRT
ORGANISM: Human
US-10-078-547-2

Query Match 9.4%; Score 134.5; DB 13; Length 503;
Best Local Similarity 25.8%; Pred. No. 0.071;
Matches 63; Conservative 23; Mismatches 107; Indels 51; Gaps 11;

QY 42 PAGTR-PTPPDPPEVKEANRTGVGKEASKALGAKGSCETTPSRGSTPT--LTPRKKX 98
DB 207 PGFPRQSPGPTPPPPGNGRTALGGGSIQSPSSSPFSNRPLPTPSALDDKPP 266
QY 99 KYRPSHTPSYCDLSFGSRSEGSFGAPRMAKGAALRLMTPPPPRSGSHSPRR- 157
DB 267 PPPPGNPPSIHREAV-----PPPPQNNKPPVPS 297
QY 158 ---EAPLPA-IHPAGSKTEPPPADSQKLSWGLHSSRLKRGLSHLTNVSTGH- 212
DB 298 PRPSAPHPRLRPPPPSPGPPPLPS---SSGNDETPLPQKNTLSSTTPPLSPGSS 354
QY 213 -PATSAPHTNGPQDLR--PSTSGVTPRSPLVTSRAVSISVSTPR--RGATQK--- 263
DB 355 GPLPVPSEBRPPVPDRDPGRSG-PLPPPPVSRNGSTRALPATPQLPSRSGVDSPPSG 413
QY 264 PKPP 267
DB 414 PRPP 417

RESULT 11
US-10-084-846A-7
Sequence 7, Application US/10084846A
Publication No. US2004006026A1
GENERAL INFORMATION:
APPLICANT: WEITNAUER, GABRIELE
APPLICANT: MUEHLEWEG, AGNES
APPLICANT: TRETZER, AXEL
APPLICANT: BECHTHOLD, ANDREAS
TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
FILE REFERENCE: 1974-005
CURRENT APPLICATION NUMBER: US/10/084,846A
CURRENT FILING DATE: 2003-02-25
PRIOR APPLICATION NUMBER: PCT/EP01/09815
PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: DE 101 09 166.4
PRIOR FILING DATE: 2001-02-25
NUMBER OF SEQ ID NOS: 120
SOFTWARE: PatentIn Ver. 3.2
SEQ ID NO 7
LENGTH: 19652
TYPE: PRT
ORGANISM: Streptomyces viridochromogenes
FEATURE:
OTHER INFORMATION: Protein 2: amino acid sequence encoded by coding strand 2.
US-10-084-846A-7

Query Match 9.4%; Score 134.5; DB 15; Length 19652;

Best Local Similarity 25.7%; Pred. No. 4.1;
Matches 80; Conservative 29; Mismatches 103; Indels 99; Gaps 16;

QY 26 RYKARTSYVDETLPFGSPG-----TRPPDPDPPEVKEANRTGVGKEASKAL 74
DB 11238 RRPVPAHVDAEAAIGLVGVADVDVPSRSVIPSELPQWVER-----ROSETPMSAS 11293
QY 75 GAKGSC-----ETTPSGSTPLTP-----RKKNKTRP--- 102
DB 11294 SSQGRCTWRSPAQSVASXWPPSPANSIDSTASRSST-LVTPCTSMTNSARRRFFSSPSBA 11352
QY 103 -ISHTPSYCDLSFGSRSEGSFGAPRMAKGAALRLMT-----TPP 145
DB 11353 RIS-TPG-CPDIAVRSAPSSAPSSSTRILAPGAANASARSGWLSRVAVAVYRPL 11410
QY 146 PTPRGSH--SPRPPAPRLAIHPAGSKTEPPPADSQKLSWGLHSSRLKRGLSHSLT 203
DB 11411 YTKSFTHGSSSTRRRRISRSFTVALP-----PLSGALPSETWAGDHRSAPLRSATYH--- 11463
QY 204 HLNVPSTGHPAATSA-----HTNGPD-----LRPSTGVTRSPPLVTSRAKSV 247
DB 11464 -----CAHSGTRSPISISKEBSSAGPTAAHAARVASPSASVSAQSP-KTISRASSV 11516
QY 248 SISVSTPRRG 258
DB 11517 AVATTSSRAG 11527

RESULT 12
US-10-437-963-143626
Sequence 143626, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 143626
LENGTH: 625
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
NAME/KEY: unsure
LOCATION: (1) (625)
OTHER INFORMATION: unsure at all Xaa locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MKT4530_44516C.1.pcp
US-10-437-963-143626

Query Match 9.3%; Score 134; DB 16; Length 625;

Best Local Similarity 23.5%; Pred. No. 0.098;
Matches 73; Conservative 27; Mismatches 103; Indels 108; Gaps 13;

QY 41 SPAGTRPT-----PDPDPPEVKEANRTGVGKEASKALGAKGSCETTPSRGSTPTL 92
DB 227 APASTAXPHPRPHRPLPXGSP-REAAARPPPLSDLAVALSPSGAAAPPR----- 279
QY 93 TPRKKNKTRPSHTPSYCDLSFGSRSEGSFGAPRMAKGAALR----- 138
DB 280 -----REVAFTPPRDSASASARPO-----PRGPAALAAATARGPOASRASP 325
QY 139 -----ALMTPPPTPRGSHS-----PPREAPRLAIRPAGPSK 171

Db 326 PRVVRPRPAQADKGIITTPPTPMQARAGPGRVVRVPPKRPTRGSSPPHLPPCR 385
Qy 172 -----TERGPADSOQLSMGSLHSLRLKGLSLTHLVNP---STGHDPATSA 217
Db 386 PGRBACILLSPSPSPASISLASVA---CPABSPSTLPKTRPPSSPPKSTABPSTBA 442
Qy 218 --PHTNGPDLRPSTSGVTRSPPLVTSRRAR-----SVSISV-----PST-PR 256
Db 443 SPPLISASISRRRSPSPSPPLVTSRRQSTPSPPPPSISLASVACCPABSPSTLPR 502
Qy 257 RGATQKPKP 267
Db 503 RTRPSSPPSP 513

RESULT 13
US-10-437-963-156445
; Sequence 156445, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT FILING DATE: 2003-05-14
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 156445
; LENGTH: 937
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(937)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_56111C.1.pcp
US-10-437-963-156445

Query Match 9.3%; Score 134; DB 16; Length 937;
Best Local Similarity 26.2%; Pred. No. 0.15;
Matches 68; Conservative 21; Mismatches 83; Indels 88; Gaps 14;

Qy 40 GSPAG-TRTPPPDFDPWEKANRTRGVGEKASKALGAKSCETTPSRGS----- 88
Db 702 GPPGPPSRPTOP---PEPLPSRVTR-----VDAAPTAPSRASPPPHILSPS 747
Qy 89 -----TTLTPRKKNKYRPISTPSYCDSESLF-GSRSEGASFGAPMAKDAKL 137
Db 748 SSLSRAAGRLQPLSLR-PSAARPCRLRRRSRLIFAGAASNSWGNHRRRLVLAGKVAL 806
Qy 138 RALLWTPPTPGSHSPRPRE-APLR---AHPAGS---KTEPPADSOQLSMG--G 187
Db 807 PPLYLQPLPRRSAGRLRLRLPRRRVALHPLGSLVLPERRRP-----RRLSLGLVG 860
Qy 188 LHSRPLKRLSHSLTHLVNPSTGHDPATSAHTNGPDLRPSTSGVTRSPPLVTSRRAR 247
Db 861 THSGSPRR-----ASAPADVAVTSIGIA-----AGKV 888

Qy 248 SISV---PSTPRRGATQKPK 264
Db 889 VLGIASPPSKRLMGSSSP 908

RESULT 14
US-10-437-963-200034

; Sequence 200034, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT FILING DATE: 2003-05-14
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
US-10-437-963-200034

Query Match 9.3%; Score 133; DB 16; Length 380;
Best Local Similarity 24.2%; Pred. No. 0.067;
Matches 62; Conservative 16; Mismatches 114; Indels 64; Gaps 9;

Qy 41 SPAGTRPPDFDPWEKANRTRG-----VGKESKALGAKSCETTPSRG 87
Db 86 SAAAAAPPAPAPPPRRRRRCGNPPPLPPRSSASASAPSRASPPAPATCPAP 145
Qy 88 STPTLTPRKKNKYRPISTPSYCDSESLFSGRSEGASFGAPR---MAKDDAKLALMT 143
Db 146 RCGSATP--PAPWRPSPSPSPAPAPACGCAARTPPAYRRLSVITPPPHHAAVL 203
Qy 144 PPPTPR---GSHSPRRAPLRAIH-----PAGSKTEPGPADSOQLSMGLHSRP 193
Db 204 PPAPPRRNTRTPRRPRRRHTLRIPPPPLPAPSPAPPG----- 246
Qy 194 LKRGSLSHLTHLVNPSTGHDPATSAHTNGP--QDLRPSGVTFRSPPLVTSRRARSISV 251
Db 247 -----RHHPPEP-HPXPGRPPRPAGRLHPSPPPPNPNPLPSSR-----GP 290
Qy 252 PSTPRRGATQKPKP 267
Db 291 PPSPSRSPATDLPAP 306

RESULT 15
US-10-437-963-111701
; Sequence 111701, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT FILING DATE: 2003-05-14
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966

```

; SEQ ID NO 111701
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(514)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_15655C.1.pap
US-10-437-963-111701

```

```

Query Match          9.3%; Score 133; DB 16; Length 514;
Best Local Similarity 26.6%; Pred. No. 0.094;
Matches 64; Conservative 20; Mismatches 117; Indels 40; Gaps 9;

```

```

QY 42 PAGTRPTPPDPPWVEKANRTRGVQKEAS-----KALGAKGSCETTPSRGSPPTLTP 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 239 PARTLPAPPXADPPGRGXPRRRATGTRAAPPFACTRRSPAGRCGCSPPRPPGSPPTPAP 298
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 95 RKKNKYRPISTPSYCDESLFGSRSEGASFGAPRMAKGDALRLALMTPTPTPRGSHSP 154
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 299 -----PPPPPPRGAPRAAPASAAAPASAAAGRAPRAPRPPRPPPPPSARASSSP 352
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 155 RPREAPLRA---IHPAGSPKTEPGPADSQKLSMGGLHSRPLKRGGLSHSLTHLNVPT 210
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 353 RHSATPAPAHKPRRPSFGHGRPPPTTRAPSS-----PWPFRITGAAP-----RGPET 398
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 211 GHPATSAPTHNG-PODLR---PSTSGVTFRSPLVTSRARSVSISVSTPRRGATOKPKP 266
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 399 PAPSSSWPPPAGTPAALRWADADASAAGTACS--AAGRAAG---SSPPPPQHNNHHPSPP 453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 267 P 267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 454 P 454
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: March 28, 2005, 09:05:30
 Job time : 391 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2005, 08:00:05 ; Search time 24 Seconds
(without alignments)
1078.430 Million cell updates/sec

Title: US-10-031-589-4

Sequence: 1 MKTVELAVSGMQLGLQHR.....SVSPTRRGATQKPKPPWK 269

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : PIR 79: *
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	144.5	10.1	539	2 T28770	hypothetical prote
2	136	9.5	2715	2 T13049	eyelid - fruit fly
3	132.5	9.2	580	2 T43481	probable mucin DKF
4	130.5	9.1	403	2 A35363	synapsin I splice
5	128	8.9	403	2 S52796	prl2 protein - hu
6	127.5	8.9	1870	2 S37671	MHC class III hist
7	127.5	8.9	1872	2 S36152	MHC class III hist
8	127	8.8	302	2 T32711	hypothetical prote
9	126.5	8.8	1199	2 A40670	nuclear envelope p
10	126	8.8	706	2 E30411	US protein - huma
11	125.5	8.7	699	2 C43674	probable, serine/th
12	124	8.6	576	2 T36728	hypothetical prote
13	123.5	8.6	724	2 T47149	glycoprotein 350/2
14	123	8.6	886	2 S29605	membrane antigen 9
15	122.5	8.5	907	1 OQBE21	hypothetical prote
16	122	8.5	1777	2 T34369	hypothetical prote
17	122	8.4	721	2 E70766	high molecular mas
18	121	8.4	1151	2 T18535	MHC class III hist
19	120.5	8.4	2142	2 B50398	NIMA-like protein
20	120	8.4	779	2 A57177	unconventional myo
21	120	8.4	3530	2 A59266	immediate-early pr
22	119	8.3	1446	1 A45344	serine-rich protei
23	118.5	8.2	1077	2 A44067	unconventional myo
24	118.5	8.1	306	2 T52305	En/Spm-like transp
25	116.5	8.1	3511	2 A59295	unconventional myo
26	116	8.1	635	2 F75477	hypothetical prote
27	115.5	8.0	846	2 T21700	hypothetical prote
28	115.5	8.0	883	2 A96662	hypothetical prote
29	115.5	8.0	1285	2 T14171	ataxin-2 - mouse

30	115.5	8.0	1456	2 T01397	lTR gag/pol polypr
31	115.5	8.0	2176	2 T13806	lucan gene protei
32	115.5	8.0	2212	2 A41098	calcium channel pr
33	115	8.0	380	2 D70516	hypothetical prote
34	115	8.0	704	2 A30411	synapsin Ia - rat
35	114.5	8.0	308	2 S77938	EBNA-LP protein -
36	114.5	8.0	1211	2 T42230	AP4 protein - mous
37	114.5	8.0	2187	2 T30826	nascent polypeptid
38	114	7.9	2346	2 T13829	Trp homolog - fruit
39	113.5	7.9	346	2 T46916	hypothetical prote
40	113	7.9	1952	2 T48814	hypothetical prote
41	112.5	7.8	2282	2 T42717	DNA-binding protei
42	112	7.8	214	2 T10737	extensin-like cell
43	112	7.8	214	2 T10737	proline-rich cell
44	112	7.8	620	2 S06733	hydroxyproline-ric
45	112	7.8	3149	1 OQBE8	Bpuf1 protein - hu

ALIGNMENTS

```
RESULT 1
T28770
hypothetical protein W03D2.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28770
R:Roßling, T.; Wohlmann, P.
submitted to the EMBL Data Library, June 1997
A:Description: The sequence of C. elegans cosmid W03D2.
A:Reference number: Z20519
A:Accession: T28770
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-539 <ROH>
A:Cross-references: UNIPROT:Q8W0G9; EMBL:AF000298; PIDN:AAC48255.1; GSPDB:GN00022; CESP:
A:Experimental source: strain Bristol N2; clone W03D2
C:Genetics:
A:Gene: CESP:W03D2.1
A:Map position: 4
A:Insertions: 40/3; 88/3; 115/3; 146/3; 173/3
C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

Query Match      10.1%; Score 144.5; DB 2; Length 539;
Best local Similarity 24.1%; Pred. No. 0.025;
Matches 75; Conservative 18; Mismatches 97; Indels 121; Gaps 13;

QY      13  QTLGLQHRGCGYVVKARTSYVDLFGSPAGTRPTPDPPWVEKANRRGVGKESK 72
          |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      234  QGFGQQQQTONGFRGKRO-----APPAGSPPPP--P-----KSGP 268
          |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY      73  ALGAKGSCETPPSRGSPPTLTPRKKNKYRPISHTPSYCSDESLFGSRSEGAFGAPRMAGK 132
          |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      269  PLAGSGPPPPPAAGSP--PPTGSPPPPPPTGSP-----PPPAAGG 308
          |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY      133  DAALKRALMTPTPTPPGSG-----HSPPR-----REAPLRAIHPA 167
          |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      309  SPPPRAGSPPPPPPPSPTGSLPPPAAGSPPPAAGTSGPPPPPKRQKRAPESSPTG 368
          |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY      168  GPSKTEP-----GPAADSOXLS-----MGSG-----LHSSRLPKRGL 198
          |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      369  SPTPGSPPTGPPPGGPKSSSESSREGRCGGPGGPGGPKSSSESSSEKREPPGP 428
          |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY      199  SHSLTHLVNPSGTGHPATSAPTNGPQDLRPSTGVTFRSPLVTSRARSVSISVSPSTRPG 258
          |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      429  RRS-----PPTGSPPTGSPPTGPPPGSGPTG-----SPTGLPSRQKQ 468
          |||  :  :  :  :  :  :  :  :  :  :  :  :  :

QY      259  GATQKP--KPP 267
          |||
DB      469  APEDRPTGSP 479
```

RESULT 2

T13049 evelid - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
A:Accession: T13049
R:Treisman, J.E.; Luk, A.; Rubin, G.M.; Heberlein, U.
submitted to the EMBL Data Library, March 1998
A:Reference number: Z17592
A:Accession: T13049
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-7715 <TR>
A:Cross-references: UNIPROT:Q81N94; EMBL:AF053091; NID:G2981220; PID:G2981221; PIDN:AACCA
C:Genetics:
A:Gene: eld
A:Cross-references: FlyBase:FBgn0003013
C:Function:
A:Description: could act as a transcription factor antagonistic to the wg pathway
C:Keywords: DNA binding

Query Match	9.5%;	Score 136;	DB 2;	Length 2715;
Best Local Similarity	22.5%;	Pred. No. 0.47;		
Matches	62;	Conservative	30;	Mismatches 119; Indels 64; Gaps 11;

Qy	23	GGIRV-----KANTSYDEFTLFGSPAGTRPRPF-----DPPVEKANKRTGVG	67
Db	632	GGYTKGGGPGSGPAGQGYRPPQOYRPPENYRPRRQGYRGATATGPRRPPTSQA-----GAG	687
Qy	68	KEASKALGAKSGCETTPRSGSTPLTLPRKKKRYRISHTPRYCDSESLGSGSEGASFGAR	127
Db	688	GANMPSGAQNG--GYRGG-----MRNMQYRYQQVYRVSRCQIYRGARGAAMQGNH	740
Qy	128	RMAGGDAKKLALLMTP-----RPTPRGSHPRPRRPAIRALINRAGFSTTERPAADS	180
Db	741	VQGGK-----TPRPVVGSGRRPRQSGSGSRPLNYLQKLNHGKYGSGSRTPRQGR	790
Qy	181	QKLSMG--GLHSRRPL-----KRGLSHSLTHLNVST-----GHAPTSAPHT	220
Db	791	QGYNGNPTGMNKGMPMBRPHNMGPRHGTNNMGRTSTSTPRQSGMLQGGQPRQSGASGGER	850
Qy	221	NGPDDLPRSTSGVTRFPLVTSRRASVSIYVSTPR	255
Db	851	GGREHI--SQDNGISSGPTGAAGMAHVAIVSVTTGPR	884

RESULT 3

Probable mucin DKFZp434C196.1 - human (fragment)
 N:Alternate names: Protein DKFZp434B0635.1
 C:Species: Homo sapiens (man)
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
 C:Accession: T13481, T34549, T17264
 R:Koehler, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, December 1999
 A:Reference number: 22514
 A:Accession: T13481
 A:Molecule type: mRNA
 A:Residues: 1-580 <AAA>
 A:Cross-references: UNIPROT:Q0NP83, EMBL:AL133561, NID:g6599133, PIDN:CAB63715.1, PID:g6599133
 A:Experimental source: adult testis; clone DKFZp434C196
 R:Poustka, A.; Wellenreuther, R.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, October 1999
 A:Reference number: 221540
 A:Accession: T134549
 A:Molecule type: mRNA
 A:Residues: 262-580 <POU1>
 A:Cross-references: EMBL:AL122069, NID:96102864, PIDN:CAB59245.2, PID:g7018420
 A:Experimental source: adult testis; clone DKFZp434B0635
 R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, September 1999
 A:Reference number: 218723
 A:Accession: T17264
 A:Molecule type: mRNA

A;Residues: 262-580 <POU2>
A;Cross-references: EMBL:AL117481; NID:g5911958; PIDN:CAB55954.1; PID:g5911959
A;Experimental source: adult testis; clone DKFZp434B061
C;Genetics:
A;Note: DKFZp434C196.1; DKFZp434B065.1; DKFZp434B061.1

Query Match	9.2%	Score 132.5	DB 2	Length 580
Best Local Similarity	26.3%	Pred. No. 0.16	Mismatches 125	Indels 61
Matches 77	Conservative 30			Gaps 17

```

Qy      3 TPVELAVSGMOTLGLQHRGGRVAKARTSY-----DETLEFGSAGTRPT--PPDFDPW 56
          || :: :: | | | | | | | | | |
Db      46 TPFRASTTKMESTAL-----LRTLEPRASIMRTPTPRASIMRTPPRASFPKRPPRASPT 90

```

```
Qy      57 VEKANRTGVGKEAKSGALGAGKSCETTPSRGSTPTLTTPRKNNKYRPISHTPSYCDESLFG 116
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      99 PSRASPTRLPRASPMGSPHRASPMTPTPPRAS-PTCTPSTASP---TGTPS--SASPTG 151
```

```
Qy      117 SRSEGASFGAPRMA----KQDAAKI-----RALLTWPPPTTGGSHSDPRPREARLRAIHDA 167
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      152 TPPRASPTGPPRAMATRSPSTASTLRTPRASASLRTWPPRASPTRTD-PRSP-RMSHRA 209
```

```

Qy      168 GPKTEP--GPAADSQLKSLMGSLHSRPLKRGL--SHSLTHLINVPTGHPATSAPTHNGP 223
      |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      210 SPTTPPRASPTRRRPRAS----PRTTPRESLRTSHRASPTRMPPRASPTRRRP----- 260

```

```

QY      224 QDLRPTSGVTFR-SPLVTSRARSVISVSPSTPRRGATQPK-----PPW 268
          | | : | | : | | | | | | | | | | | | | | | | | |
Db      261 ---RASPTGSPPRASPMTPPRAS-----PRTPPRASPTTTPSRASLRTTPSW 304

```

RESULT 4

synapsin I splice form a - human
 C:Species: Homo sapiens (man)
 C:Date: 17-Aug-1990 #sequence_revision 17-Aug-1990 #text_change 09-Jul-2004
 C:Accession: A35563; B53563; A35805
 R:Stedhof, T.C.
 U: Biol. Chem. 265, 7849-7852, 1990
 U:Title: The structure of the human synapsin I gene and protein.

A1:Accession: A35363
A1:Molecule type: DNA
A1:Residues: 1-705 <SUB>

A:Accession: B35363
A:Status: preliminary
A:Molecule type: DNA
GB:U05343

A; residues: 1-657; MSFPAQF <50>
A; Cross-references: GB:M837; GB:J05431
R.Sauerwald, A.; Hoesche, C.; Oeschwald, R.; Kilmann, M.W
T. Biol. Chem. 265, 14932-14937, 1990

A: Molecule type: DNA
A: Reference number: A35805; MUID: 903686667; PMID: 2118519
A: Title: The 5'-flanking region of the synapsin I gene. A G+C-rich, IVa- and CAAT-less
A: Accession: A35805

A:Accession: 1-125 <SAU>
A:Cross-references: GB:M55301; NID:G3338655; PID:AA60608.1; GB:J056330
C:Genetics:
A:Gene: GDB:SYN1

A;Cross-references: GDB:119606; OMIM:313440
A;Map position: Xp11.23-Xp11.23
C;Keywords: actin binding; alternative splicing; phosphoprotein

Query Match	9.1%;	Score 130.5;	DB 2;	length 705;
Best Local Similarity	23.5%;	Pred. No. 0.27;		
Matches	61;	Conservative	24;	Mismatches 108;
				Indels 67;
				Gaps 9;

[illegible]

Db 512 TSAPO-----OPASQAAPPTGCGRQSRPVAGPGAPAPAPAPSPSPORQAGPPQATRQ 566
Qy 150 -----GSHSPRRERAPLRAIHAPGPKTRPGPAADSOQLSMGSLHSRLKGLSLTHL 205
Db 567 TSVSGPAPPKASGAPPGGQROGPPQKPPGAPGPPROASQAG----- 608
Qy 206 NVPSGTGHATAPHTNGP-----ODLRPSTSGVTPRSP---LVTSRASV 247
Db 609 PVPRGTGPTTQQPRPSGCGRPAAGAPKPKQLAQKPSQDVPPPPATPAAGPPHPLQNSQSILTN 668
Qy 248 SISVSTPRRGATQKPP 267
Db 669 AFNLP-----EPAP 678

RESULT 5
S52796
Prp2 protein - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 19-May-1995 #sequence_revision 10-Nov-1995 #text_change 10-Sep-1997
C/Accession: S52796
R/Ruhlmann, A.; Kreideweis, S.; Nordheim, A.
Submitted to the EMBL Data Library, March 1995
A/Reference number: S52796
A/Accession: S52796
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-403 <RUI>
A/Cross-references: EMBL:X86019; NID:G762950; PID:G762951

Query Match 8.9%; Score 128; DB 2; Length 403;
Best Local Similarity 25.3%; Pred. No. 0.22; Indels 52; Gaps 11;
Matches 62; Conservative 23; Mismatches 108;
Qy 42 PACTR-PTPPDPPEWEKANRTGKAGKALGAKSCETPSRGSTP--LTPRKN 98
Db 116 PGGRRQSPGPTPPPPGNRGTALGGGSIROSPLSSSPFNRPPLPTPRBALDDKPP 175
Qy 99 KYRPISTPSYCDLSFGSRSEGASFGAPMAKGDAAKLALMLTTPPTPGSHSPRR- 157
Db 176 PPPVGNRPSTHRAV-----PPPPQNNKPPVPST 206
Qy 158 ---EAPLRA-IHPAGSKTEPGPAADSOQLSMGSLHSRLKGLSHLTHLNPSTGHP 213
Db 207 PRBAPHRPLRPPPSPPPLPPS---SSGDETPLRQKMLSLSSSTPPSPERS 263
Qy 214 A---TSAHTNGPODLR--PSTSGVTPRSPVTSRARSISVSTPR---RGATQK-- 263
Db 264 GPLPPVPSEPRPPVRPPRSG-PLPPPPVSRNNGSTSRALPATPLPSPRSQVDSPRS 322
Qy 264 -PKPP 267
Db 323 GPRPP 327

RESULT 6
S37671
MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - human
C/Species: Homo sapiens (man)
C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 15-Sep-2000
C/Accession: S37671
R/Bougueleret, L.
Submitted to the EMBL Data Library, August 1992
A/Reference number: S37671
A/Accession: S37671
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1870 <BOU>
A/Cross-references: EMBL:Z15025; NID:G29374; PID:G29375
C/Genetics:
A/Map position: 6p21.3
A/Intons: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1; 65

C:Superfamily: collagen alpha 1(IV) chain

Query Match 8.9%; Score 127.5; DB 2; Length 1870;
Best Local Similarity 25.2%; Pred. No. 1.1; Indels 73; Gaps 12;
Matches 68; Conservative 17; Mismatches 112;
Qy 33 YVDETLF-GSPAGTRPPDPDPV-----EKANTRGVGKE-----ASKALGAGS 79
Db 728 YVDETLFGR-----PLDYPYGVHPSGLVPRERSRLSSEPPRHAAPMLRENGT 781
Qy 80 CETTPSR---GSTPLTPRKKNKTRPI-SHTPSYCDLSFGSRSEGASFGAPMAKGDAA 135
Db 782 PPVDPKLAWGDVFTATPAEP---RPLTSPRLQAADDDKMRSE----- 823
Qy 136 KLRALLWTPPTPGSHSPRRERAPLRAIHAPGSKTEPGPAADSOQLSMG----- 187
Db 824 -----TPPV-----PP--PYLASYGFPENGTPGPISRFPLEEGPRLPMP 867
Qy 188 -----LHSSRPKRGSLSHLTHLNPSTGHPATSAHTNGPODLRPSGTVPSPPLV 240
Db 868 GSDEVAKIQTPPKPKPKETAGLTGPRAGKPLPASHSGAGPPPPRRESKTRTWGPR 927
Qy 241 TSRAASVSIYP-STPRRGATQKPPWK 269
Db 928 GSSRRGIPPEEPGAPPRRAGPIKKPPPTK 957

RESULT 7

S36152
MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - human
C/Species: Homo sapiens (man)
C/Date: 06-Jun-1995 #sequence_revision 17-Nov-1995 #text_change 15-Sep-2000
C/Accession: S36152
R/Iriri, F.U.M.; Bougueleret, L.; Prieur, S.; Caterina, D.; Primaes, G.; Perrot, V.; Jurka
Nature Genet. 3, 137-145, 1993
A/Title: Dense Alu clustering and a potential new member of the NFkapab family within a
A/Reference number: S36152; MUID:93272029; PMID:8499947
A/Accession: S36152
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1872 <RI>
A/Cross-references: EMBL:Z15025
A/Note: in the authors' translation residues 32-34 are shown after residue 4 and, conseq
C/Genetics:
A/Intons: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1; 65
C:Superfamily: collagen alpha 1(IV) chain

Query Match 8.9%; Score 127.5; DB 2; Length 1872;
Best Local Similarity 25.2%; Pred. No. 1.1; Indels 73; Gaps 12;
Matches 68; Conservative 17; Mismatches 112;
Qy 33 YVDETLF-GSPAGTRPPDPDPV-----EKANTRGVGKE-----ASKALGAGS 79
Db 729 YVDETLFGR-----PLDYPYGVHPSGLVPRERSRLSSEPPRHAAPMLRENGT 782
Qy 80 CETTPSR---GSTPLTPRKKNKTRPI-SHTPSYCDLSFGSRSEGASFGAPMAKGDAA 135
Db 783 PPVDPKLAWGDVFTATPAEP---RPLTSPRLQAADDDKMRSE----- 824
Qy 136 KLRALLWTPPTPGSHSPRRERAPLRAIHAPGSKTEPGPAADSOQLSMG----- 187
Db 825 -----TPPV-----PP--PYLASYGFPENGTPGPISRFPLEEGPRLPMP 868
Qy 188 -----LHSSRPKRGSLSHLTHLNPSTGHPATSAHTNGPODLRPSGTVPSPPLV 240
Db 869 GSDEVAKIQTPPKPKPKETAGLTGPRAGKPLPASHSGAGPPPPRRESKTRTWGPR 928
Qy 241 TSRAASVSIYP-STPRRGATQKPPWK 269
Db 929 GSSRRGIPPEEPGAPPRRAGPIKKPPPTK 958

RESULT 8

T32711
 Hypothetical protein T22D1.2 - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
 C/Accession: T32711
 R/Giesel, C.; Bradshaw, H.; Hawkins, M.
 submitted to the EMBL Data Library, December 1997
 A/Description: The sequence of C. elegans cosmid T22D1.
 A/Reference number: Z21211
 A/Accession: T32711
 A/Status: preliminary; translated from GB/EMBL/DDBJ
 A/Molecule type: DNA
 A/Residues: 1-302 <GB>
 A/Cross-references: UNIPROT:Q9GZH1; EMBL:AF039052; PIRN:AAB94271.1; GSPDB:GN00022; CESP:
 A/Experimental source: strain Bristol N2; clone T22D1
 C/Genetics:
 A/Gene: CESP:T22D1.2
 A/Map position: 4
 A/Introns: 89/1
 C/Superfamily: proline-rich protein

Query Match 8.8%; Score 127; DB 2; Length 302;
 Best Local Similarity 24.6%; Pred. No. 0.19;

Matches 65; Conservative 15; Mismatches 98; Indels 86; Gaps 11;

QY 44 GTRPTPDPPPP-WVKAKRTR-----GVGKASKALGAKG--SCETTPRGSTPTL 92
 DB 72 GTRPPPTGPPGPDLSAEGNARRPPPPPKGTGTPPPPTGEPDLSGEGNARRPP-- 129
 QY 93 TPRKKNKVRPISTHTPSYCDLSLFGSRSEGASFGAPRAKGDAAKLRALLWTPPTPGSH 152
 DB 130 -PPKGTSSPPPPPT-----GEPDLSGEGNARR--RPPPPPKGTG 166
 QY 153 SPSP-----REAPLRAHPAGSKTEPPPADSQTLNMGGLHSSPLK 195
 DB 167 SPPPPTGEPDLSGEGNARRPP--PPKGTGTPPPPTGEPDLSAEGVARRP-- 220
 QY 196 RGLSHSLTHLVNSTGHPAISAPHTNGPDLR-----PSTGVTFRSLVTSRAR 245
 DB 221 -----PPPKGSGSPPTGTPGPDLSGEGNARRPPPPPKGTGSPPPPTGEPQ 270
 QY 246 SVSISVSTPRRGATOKPKPPMK 269
 DB 271 DLS-----GEGNARRPPPPK 287

RESULT 9

A40670
 nuclear envelope protein POW 121 - rat
 C/Species: Rattus norvegicus (Norway rat)
 C/Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
 C/Accession: A40670
 R/Haliberg, E.; Wozniak, R.W.; Blobel, G.
 J. Cell Biol. 122, 513-521, 1993
 A/Title: An integral membrane protein of the pore membrane domain of the nuclear envelope
 A/Reference number: A40670; MUID:93328754; PMID:8335683
 A/Accession: A40670
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-1199 <HAL>
 A/Cross-references: UNIPROT:P52591; GB:Z21513; NID:G396746; PIRN:CAA7925.1; PID:G396747
 F/803-807,845-849,956-960,1010-1014,1047-1051,1076-1080/Region: pentapeptide motif (X-F-

Query Match 8.8%; Score 126.5; DB 2; Length 1199;
 Best Local Similarity 26.1%; Pred. No. 0.84;

Matches 67; Conservative 26; Mismatches 131; Indels 33; Gaps 11;

QY 28 KARSYVDLTF-----GSPAGTRPTPDPPPWVKAKRTRGVGKASKALGAKGSCETT 83
 DB 362 RSRRTSVSLTSTGTGSPSSRNA-----ITSSYSTRTGV-SQLMKSGPTSSPPSS 413
 QY 84 P--SRGSTPTLTPRKKNKVRPISTHTPSYCDLSLFGSRSEGASFGAPRAKGDAAKLRALL 141

DB 414 PASSRSQTEP-RPAKTRREBPCHQSSSAPLYTDKESPEKVTDPATKQSS-----L 466
 QY 142 WTPPPPTGSHSRREAPL-----RAIHAGSKTEPPGAASOKLSMGGLHSSRLKRG 197
 DB 467 WTSPPPT-GSSGGRKKRIIDLPERRDQUTLPPPLGYSTIAEDIDMERASLQWFNKV 525
 QY 198 LSHSLTHLVNSTG-HPATSAPHTNGPDLRPSTGVT-----RSPLVTSRAVSISVP 252
 DB 526 LEPKTDASTPAIDTSPATSPPTTLTLPIVGPAAASALPAPSNLILSLKMGSSPAP 585
 QY 253 ST--PRRGATOKPKP 267
 DB 586 SSSEPEEATVAAPSDP 602

RESULT 10

E30411
 synapsin Ia - bovine
 C/Species: Bos primigenius taurus (cattle)
 C/Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 01-Dec-2000
 C/Accession: E30411; F30411; A35758
 R/Suedhof, T.C.; Czernik, A.J.; Kao, H.T.; Takei, K.; Johnston, P.A.; Horiuchi, A.; Kanai
 Science 245, 1474-1480, 1999
 A/Title: Synapsins: mosaics of shared and individual domains in a family of synaptic vesicle
 A/Reference number: A30411; MUID:8938265; PMID:2506542

A/Accession: E30411
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-706 <SUE>
 A/Accession: F30411
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-661, 'KA', 664, 'PAQAP', <SUD>
 R/Hall, F.L.; Mitchell, J.P.; Vulliamt, F.R.
 J. Biol. Chem. 265, 6944-6948, 1990
 A/Title: Phosphorylation of synapsin I at a novel site by proline-directed protein kinase
 A/Reference number: A35758; MUID:90216728; PMID:2108963
 A/Accession: A35758
 A/Molecule type: protein
 A/Residues: 532-556 <HAL>
 C/Keywords: actin binding; alternative splicing; phosphoprotein
 F/551/Binding site: phosphate (Ser) (covalent) (by proline-directed kinase) #status exper

Query Match 8.8%; Score 126; DB 2; Length 706;
 Best Local Similarity 22.9%; Pred. No. 0.52;

Matches 60; Conservative 23; Mismatches 87; Indels 92; Gaps 12;

QY 40 GSPAGTRPTPDPPPWVKAKRTRG-----VGEKASKALGAKGSCETTPSRGSTP 90
 DB 454 GPPAQGR-PPGCGPPQPGPQGRPLQGRTPGQGHSLGPPAG-SPLFRLPSP 511
 QY 91 TLTPRKKNKVRPISTHTPSYCDLSLFGSRSEGASFGAPRAKGDAAKLRALLWTP--PPTP 148
 DB 512 TSVPO-----QPASQ-----TTPMTQGGRGSRPVPAGGPAPPAT 546
 QY 149 RGSHPSPR-----EAPLRA--IHPA-----GPKTEPPADSOGLSM 185
 DB 547 RPPASPPQQAQPPQATRTQTSVGAHPPASGVPPGQGRQPPQKPPPPAPPTQAQSQ 606
 QY 186 GGLHSSRLKRLSHSLTHLVNSTGHPATSAPHTNGPDLRPSTGVTFRSLVTSRAR 245
 DB 607 AG-----PMERTGPTTQGRPSGPGAPGAPTPKQLAQKP----- 641
 QY 246 SVSISVSTPRRGATOKPKP 267
 DB 642 SDVPPPTAAAG-----PP 657

RESULT 11

C43674
 US4 protein - human herpesvirus 2 (strain HG52)
 C/Species: human herpesvirus 2

C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C:Accession: C43674
R:McGeoch, D. J.; Moss, H. W. M.; McNab, D.; Frame, M. C.
J. Gen. Virol. 68, 19-38, 1987
A>Title: DNA sequence and genetic content of the HindIII 1 region in the short unique ccc
A:Reference number: C43674
A:Accession: C43674
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-699 <MCG>
C:Cross-references: UNIPROT:P13290; EMBL:X04798; NID:959900; PIDN:CAA28490.1; PID:959904
C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

Query Match 8.7%; Score 125.5; DB 2; Length 699;
Best Local Similarity 23.0%; Pred. No. 0.56;
Matches 59; Conservative 27; Mismatches 92; Indels 79; Gaps 11;

35 DDTLFFSPACTRTPPPDDPPWVEKANTREVGEKASAKLAKGSCETTPSGSTPTLP 94
|||
405 EETAVASPPATASVE---SSPLPAAATAATPAGHTNTSSASAKPTPTPA---PTPP 457
|||
95 RKKNKRPISHTPSYCDSELFSGSRSEGSFGAPRAKGDAAKLRALLMTTP---PTPRG 150
|||
458 PRTSHATPRPTT-----GPGV-----TPPGATGPGV 486
|||
151 SHSPREAPRLRIHPAGPKTEPGPAADSQLSMGGLHSSRLKGLSHS-----LTHL 205
|||
487 ASAAPPADSPPLTASPPA---TAPGSAANVSA---ATTATPGRTGARTPTDPTKTP 539
|||
206 NVPSTGHPATSAP---HTNGPDLR-----PSTGVTRRSPPLVTSRAKSVSIS 250
|||
540 HEPADAPPGSPAPPPEPHRGGEFEEBGAGDEPPEDDDSATGLAFTPTPNK----- 591
|||
251 VESTPRGATOKPKPP 267
|||
592 --PPPARPGIRPTLP 606
|||

RESULT 12
T36729
Probable serine/threonine-specific protein kinase (EC 2.7.1.-) - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 16-Aug-2004
C:Accession: T36729
R:Murphy, L.; Harris, D.; Bentley, S. D.; Parthill, J.; Barrell, B. G.; Rajandream, M. A.
submitted to the EMBL Data Library, June 1999
A:Reference number: Z21612
A:Accession: T36729
A:Status: translated from GB/EMBL/DDAY
A:Molecule type: DNA
A:Residues: 1-576 <MUR>
A:Cross-references: UNIPROT:Q9YAK4; EMBL:AL079308; PIDN:CA45227.1; GSPDB:GN00070; SCOPED
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOPDB:SCHE9.30
C:Superfamily: protein kinase homology
C:Key words: phosphotransferase; serine/threonine-specific protein kinase

Query Match 8.6%; Score 124; DB 2; Length 576;
Best Local Similarity 26.8%; Pred. No. 0.57;
Matches 74; Conservative 20; Mismatches 90; Indels 92; Gaps 19;

41 SPAGTRPPDPDPFW--VEKANTRGVGKASAKLAKGSCETTPSGSTPTLP 98
|||
319 SPPTATPTTPP-APPQGTTPAGSSGLDRSP---GSPGPTTPD-STPASPP---- 368
|||
99 KYRPISTHPSYCDSELFSGSRSEGSFGAPRAKGDAAKLRALLMTTPPTPRGSHSPPRE 158
|||
369 PGPIVATGCPSPAPGLPPASDQG-----WT-PTSPSGPTA--PPS 405
|||
159 APRLRIHPAGPKTEPGPAADSQLSMGGLHSSR-PL-KRGLSHSLT--HINVD--- 209
|||

Db 406 AHSAPSPAPGCTRRAP-----HGHSEEVPLAERGMSETSGFHLPPQPTVT 456

Qy 210 -TGHPRTS-----APH---TNGPQDLRPST-----GVTFRSPLVTSR 243

Db 457 PTSDAASDAAMAAQBPAPHPAFATGSGRGRLPPDRAPGRSGHPAPHPGLTARS--LAPSP 515

Qy 244 ARSVSI-----SVSPTRRG-ATQKPRP 267

Db 516 ARRADVPTAAVYTAARNPRSPAPPAQHGARRRRRPSP 551

RESULT 13

T47149

hypothetical protein DKFZp547P103.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 16-Aug-2004

C:Accession: T47149, E56695

R:Blocker, H.; Boeher, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S.

submitted to the Protein Sequence Database, March 2000

A:Reference number: Z24376

A:Accession: T47149

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-724 <AA>

A:Cross-references: UNIPROT:Q04727; EMBL:AL162059

A:Experimental source: fetal brain; clone DKFZp547P103

R:Stifani, S.; Blumheller, C.M.; Redhead, N.J.; Hall, R.E.; Artavanis-Tsakonas, S.

Nature Genet. 2, 119-127, 1992

A:Title: Human homolog of a Drosophila enhancer of split gene product define a novel fa

A:Reference number: A56695, MIMD:9326515; PMID:1303260

A:Accession: E56695

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 277-418, 'P', 420-459, 'A', 461-724 <ST>

A:Cross-references: GB:M99439; NID:g307515; PION:AAA61195.1; PID:g307516

C:Genetics:

A:Gene: GDB:TLB4; BSG

A:Cross-references: GDB:228050

A:Note: DKFZp547P103.1

C:Superfamily: WD repeat homology

C:Keywords: nucleus

F:480-513/Domain: WD repeat homology <WD1>

F:566-599/Domain: WD repeat homology <WD2>

F:648-681/Domain: WD repeat homology <WD3>

F:689-722/Domain: WD repeat homology <WD4>

Query Match 8.6%; Score 123.5; DB 2; Length 724;

Best Local Similarity 23.2%; Pred. No. 0.78;

Matches 67; Conservative 38; Mismatches 113; Indels 71; Gaps 13;

Qy 3 TEVEL-AVSGMOTGLGHRCRGGYRVKARTSYVDELTFGSPAGTRPPDPFPWVEKAN 61

Db 71 TMAELNATIGQOOLAOHISHG-----HGLPVLTPRPSGLQPPALPIPG 115

Qy 62 RTRGVGEKSAKLGAKGSC-----ETPSRGSTPTLTPRKKNYRPIIS 104

Db 116 SSAGL-LALSSALGGSHLPKDEKKHNDHQDRDSIKSSVSPSASFGAEKGR--- 171

Qy 105 HTPSYCDELTFGSRSGAGFAPRMAGDAKALALMT---PPTPRGS--HSFR--- 155

Db 172 NSADYSSEK-KQTEKEETIARYDDSGESDNLVVDVSNEDSSSPGSPASPRENGL 230

Qy 156 -----PRAPPLRAHPAGSKTEPGPAADSOKLSMGLHSRPLKRLSHSLTHLVPS 209

Db 231 DKTRLKQDAP---ISPASIASSSSTPSSKSKELSLN-----EKSTPYVSKSN 275

Qy 210 TGHPRTS--PTNGPQDLRPSTSGVTFRSPLVTSRAASVSYVP-STP 255

Db 276 TPTPTDAPTPGSNSTPGLRPVPGKPPGVDPPLASLRTPMVAVPCPYTP 324

RESULT 14

S29605

glycoprotein 350/220 - human herpesvirus 4
 C,Species: human herpesvirus 4, Epstein-Barr virus
 C,Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
 C,Accession: S29605
 R,Klein, K.; Mueller-Lantzsch, N.
 submitted to the EMBL Data Library, October 1992
 A,Description: Sequences of the membrane proteins gp 350/220 and p140 of Epstein-Barr vi
 A,Reference number: S29605
 A,Accession: S29605
 A,Status: preliminary
 A,Molecule type: DNA
 A,Residues: 1-886 <KLE>
 A,Cross-references: UNIPROT:Q07284; EMBL:X67776; NID:G59163; PIDN:CAA2485.1; PID:G59164
 C,Superfamily: Epstein-Barr virus membrane antigen gp350
 C,Keywords: glycoprotein

Query Match 8.6%; Score 123; DB 2; Length 886;

Best Local Similarity 23.9%; Pred. No. 1; Mismatches 120; Indels 102; Gaps 15;

Matches 79; Conservative 30; Mismatches 120; Indels 102; Gaps 15;

```

QY 3 TPVELAVSGMOTLGLQHRCRGGRVYKARTSYVDLFGSPAGTRPT-----PDDEPPW 56
DB 529 TPTPNATS--PTLG-----KTSPTSAVTTPTPNATSPPTAVTTPTPNATPT 573
QY 57 VEKANRTG-----VKEASKA-----LGAKSGCE--TTPSRGSTPTLTPRK 96
DB 574 LGKTSPTSAVTTPTPNATSPVTGGETSPQANTNHTLGGTSTPVVTSPPKNAATSAVTTGQ 633
QY 97 KN-----KRPISHTPSYCDSEILFGSRSEGASFGAPRMAKG-----DAAKLRAL 140
DB 634 HNTSSSTSSMSLRPSSISFTLSPTSDNSTSHMPLTSAHPGGENITQVTPRSTTH 693
QY 141 LMTPEPTPT-----GSHSPRPREA-----PLRAIHPAGS-----KTE 173
DB 694 VSTSSPARPPTQSGASPGNSSTKPGEVAVTKGTPPKNATSPQAPSGKTAVPVTS 753
QY 174 PGPAADSOGLSMGILHSRPLKRGLSHSLTHLNPST---GHPATSAHPHTNGQDLRPST 230
DB 754 TGGKANS---TTGGKHTT-----GHGARTSTETPTDYGDSSTTPRTRYNATYLLPST 803
QY 231 SGV-----TFRSPLVTSRARSIVSPSTPR 256
DB 804 SKLRPRWTFTSPPTTAQATVPVPPTSQPR 834

```

RESULT 15

OQBE21

membrane antigen gp350 - human herpesvirus 4 (strain B95-8)

C,Species: human herpesvirus 4, Epstein-Barr virus

C,Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 09-Jul-2004

C,Accession: A43042; S33008; S33009; A03762

R,Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.

Mol. Biol. Med. 1, 21-45, 1983

A,Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus

A,Reference number: A93065; MUID:85035713; PMID:6092825

A,Accession: A43042

A,Molecule type: DNA

A,Residues: 1-907 <BAN>

A,Cross-references: UNIPROT:P03200; GB:V01555; GB:J02070; GB:K01729; GB:K01730; GB:V0155

R,Barrell, P.J.

submitted to the EMBL Data Library, March 1988

A,Reference number: S32973

A,Accession: S33008

A,Status: preliminary

A,Molecule type: DNA

A,Residues: 1-907 <PAR>

A,Cross-references: EMBL:V01555; NID:G59074; PIDN:CAA2485.1; PID:G1334869

A,Accession: S33009

A,Status: preliminary

A,Molecule type: DNA

A,Residues: 1-907 <PA2>

A,Cross-references: EMBL:V01555; NID:G59074; PIDN:CAA2485.1; PID:G1334869

R,Barrell, P.J.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H

Nature 310, 207-211, 1984

A,Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.

A,Reference number: A03794; MUID:84270667; PMID:6087149

A,Contents: annotation; protein coding region

C,Superfamily: Epstein-Barr virus membrane antigen gp350

Query Match 8.5%; Score 122.5; DB 1; Length 907;

Best Local Similarity 22.8%; Pred. No. 1.1; Mismatches 128; Indels 109; Gaps 13;

Matches 77; Conservative 24; Mismatches 128; Indels 109; Gaps 13;

```

QY 3 TPVELAVSGMOTLGLQHRCRGGRVYKARTSYVDLFGSPAGTRPT-----PDDEPPW 56
DB 543 TPTPNATS--PTLG-----KTSPTSAVTTPTPNATSPPTAVTTPTPNATPT 573
QY 50 PPDEPPWEKANRTGVEKASKALG-----AKSGCETTPERGSTPTLTPRKQKY 100
DB 588 PNATSPTLGKTSPTSAVTTPTPNATSPVTGGETSPQANTNHTLGGTSTPVVTSPPKNAAT 647
QY 101 RPI-----SHTPSYCDSEILFGSRSEGASFGAPRMAKGDAAKLRALLMTTP 145
DB 648 SAVTTGQHNTSSSTSSMSLRPSSISFTLSPTSDNSTSHMPLTSAHPGGENITQVTP 707
QY 146 -----PTPR-----GSHSPRPREA-----PLRAIHPAGS----- 170
DB 708 ASISTHIVSTSSPARPPTQSGASPGNSSTKPGEVAVTKGTPPKNATSPQAPSGKOT 767
QY 171 -----KTEPPAADSOGLSMGILHSRPLKRGLSHSLTHLNPST---GHPATSAHPHTNGP 223
DB 768 AVPTVTSSTGKANS---TTGGKHTT-----GHGARTSTETPTDYGDSSTTPRTRYNAT 817
QY 224 ODLPSTSGV-----TFRSPLVTSRARSIVSPSTPR 256
DB 818 TYLPSTSSKLRPRWTFTSPPTTAQATVPVPPTSQPR 855

```

Search completed: March 28, 2005, 08:53:43

Job time: 28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: March 28, 2005, 05:59:19 ; Search time 80 Seconds
(without alignments)
1721.866 Million cell updates/sec

Title: US-10-031-589-4

Perfect score: 1437
Sequence: 1 MKTEVELAVSGMQTLGLQHR.....SVSPRRRGATKQKPKPMWK 269

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 03:*
1: uniprot_sprot:*
2: uniprot_crembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1437	100.0	269	2	Q96K30
2	867.5	60.4	253	2	Q9P1H0
3	863	60.1	177	2	Q96K25
4	150	10.4	647	2	Q8S148
5	144.5	10.1	518	2	Q8MG08
6	144.5	10.1	524	2	Q02123
7	144.5	10.1	539	2	Q8MG09
8	144	10.0	1067	2	Q6NV05
9	142.5	9.9	964	2	Q26963
10	140	9.7	415	1	SVNL CANPA
11	138.5	9.6	659	2	Q6C708
12	136.5	9.5	2185	2	Q75M06
13	136.5	9.5	4911	1	MIL3 HUMAN
14	136	9.5	769	2	Q9L171
15	136	9.5	2556	2	Q7K5E8
16	136	9.5	2716	1	OSA DROME
17	134.5	9.4	503	1	WALP HUMAN
18	134	9.3	463	2	Q6VCS
19	134	9.3	1131	2	Q9VLM2
20	133.5	9.3	1303	2	Q78Y42
21	133	9.3	862	1	MCUL RAT
22	133	9.3	1240	2	Q9DMH8
23	132.5	9.2	488	2	Q87W07
24	132.5	9.2	506	2	Q69125
25	132.5	9.2	580	2	Q9UFB3
26	132.5	9.2	3288	2	Q75SD9
27	132	9.2	356	2	Q9Q0B7
28	132	9.2	356	2	Q9Q0B8
29	132	9.2	1374	2	Q6LAA1
30	131	9.1	3326	2	Q75591
31	130.5	9.1	450	2	Q6MYW6

32	130.5	9.1	510	2	Q6WZU9	Q6mz9 homo sapien
33	130.5	9.1	705	1	SVNL HUMAN	P17600 homo sapien
34	130.5	9.1	917	2	Q873D8	Q873d8 neopspora
35	130	9.0	546	2	Q7P212	Q7p212 anopheles
36	130	9.0	1725	2	Q621F2	Q621f2 burkholderi
37	129.5	9.0	506	2	Q8AZK7	Q8azk7 human herpe
38	129.5	9.0	863	1	MIL3 HUMAN	Q8m3f8 homo sapien
39	129.5	9.0	890	2	Q8UZE1	Q8uzel cercopithec
40	129.5	9.0	1014	2	Q9PUJ8	Q9puj8 oryza sativ
41	129.5	9.0	1196	2	Q6NRV6	Q6nrv6 xenopus lae
42	129	9.0	356	2	Q9Q0B5	Q9q0b5 human herpe
43	128.5	8.9	870	1	MIL3 MOUSE	Q8bqf6 mus musculi
44	128	8.9	356	2	Q9Q0B2	Q9q0b2 human herpe
45	128	8.9	356	2	Q9Q0B3	Q9q0b3 human herpe

ALIGNMENTS

RESULT 1
Q96K30 PRELIMINARY; PRT; 269 AA.
ID Q96K30;
AC Q96K30;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Hypothetical protein FLJ14627 (Hypothetical protein PSEC0043).
GN Name=FLJ14627;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obaishi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsura N., Sato K., Tanikawa M., Yamazaki M.,
RA Niimiya H., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togojya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Motekawa S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujitawa T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta S., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togoishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs";
RI Nat. Genet. 36:40-45(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosek S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,
 RA Vallatton D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Boulford G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butcherfield Y.S.,
 RA Kzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RA Strausberg R.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN SEQUENCE FROM N.A.
 RA Oca T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayaashi K., Ishii S.,
 RA Saito K., Yamamoto J., Makamatsu A., Nagai T., Nakamura Y.,
 RA Nagahara K., Sugano S., Isegai T.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK027733; BAB55328.1; -
 DR EMBL; BC022092; AAH22092.1; -
 DR EMBL; AK075358; BAC11568.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 269 AA; 28619 MW; 3F053E1454F60773 CRC64;
 Query Match 100.0%; Score 1437; DB 2; Length 269;
 Best Local Similarity 100.0%; Pred. No. 2,8e-81;
 Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKTPVELAVSGMOTLGLQHRCRGGRYKARTSYVDLTFSPSPAGTRTPPPDPPEWYKA 60
 DB 1 MKTPVELAVSGMOTLGLQHRCRGGRYKARTSYVDLTFSPSPAGTRTPPPDPPEWYKA 60
 QY 61 NRTRGVCKEASKALGANGCETTPSRGSPPTLTPRKKNKRYPSHTPSYCDSEILFSGRSR 120
 DB 61 NRTRGVCKEASKALGANGCETTPSRGSPPTLTPRKKNKRYPSHTPSYCDSEILFSGRSR 120
 QY 121 GASFGAPRMKGDMAKLRALMTPTTPPPRSGSHSPRPREALRAIHPAGSKTEPGPAADS 180
 DB 121 GASFGAPRMKGDMAKLRALMTPTTPPPRSGSHSPRPREALRAIHPAGSKTEPGPAADS 180
 QY 181 OKLSMGLHSSRLPKRGLSHSLTHLNPSTGHPATSNPHNGODLRPSTGVTFRSPPLY 240
 DB 181 OKLSMGLHSSRLPKRGLSHSLTHLNPSTGHPATSNPHNGODLRPSTGVTFRSPPLY 240
 QY 241 TSPARSVSISVPTPRRGATQKPKPPWK 269
 DB 241 TSPARSVSISVPTPRRGATQKPKPPWK 269
 RESULT 2
 Q9DIH0 PRELIMINARY; PRT; 253 AA.
 ID Q9DIH0
 AC Q9DIH0;
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
 DT 25-OCT-2004 (T-EMBLrel. 28, Last annotation update)
 DE Mus musculus 18-day embryo whole body cDNA, RIKEN full-length enriched
 DE library, clone:1110008J03 product:hypothetical protein, full insert
 DE sequence (RIKEN cDNA 1110008J03).
 GN Name=1110008J03Rik;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RX MEDLINE=92279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:119-44(1999).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RA The FANTOM Consortium.
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RX MEDLINE=2049374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Komno H., Okazaki Y., Muramatsu M., Hayaashizaki Y.,
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RN SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Komno H., Akiyama J., Nishi K., Kitsuaki T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto K., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,
 RA Onozaki Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayaashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RN SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hirooka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai J., Kojima Y., Komno H., Kouda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Nunazaki R., Ohno M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai T.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toyota T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayaashizaki Y.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnae.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Berger J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loguelli N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Rane J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalske U., Smailus D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Maita M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [8]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;
 RL Straube R.;
 DR Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK003581; BAB22871.1; -
 DR EMBL; BC021365; AAH21365.1; -
 DR MGD; MGI:1922021; 1110060303Rik.
 KM Hypothetical protein.
 SQ SEQUENCE 253 AA; 27070 MW; 239EC9BD9CFAFA61 CRC64;

Query Match 60.4%; Score 867.5; DB 2; Length 253;
 Best Local Similarity 66.3%; Pred. No. 3.6e-46;
 Matches 171; Conservative 15; Mismatches 67; Indels 5; Gaps 1;

QY 12 MQTGLQHRGCRGVKARTSYVDETLFGSPAGTRPPDPDPVWEKANTRGVGEAS 71
 DB 1 MKALHLQHRSPSTYKVKARASVDETLFGSPARTPAQDPDPVWQCNBRGVPGP 60
 QY 72 KALGKSCCTTPSGSTPTLPRKKNKRPISHTPSYCDSELRSSBEGASFGAPRAK 131
 DB 61 KGLAKRCESPSGSPNLTLPKKNKRLIGHAPSYCDESLFGTSKESG-----RAV 115
 QY 132 GDAALRALALMTPTPTPGSHSPREAPLRAIHAPGSKTEPPGAADSOKLMSGGLSS 191
 DB 116 GDAKLRLTLFTPTPTPGSHSPREAPLRAIHAPGSKTEPPGAADSOKLMSGGLSS 175
 QY 192 RPLKGLSHSLTLHNVSTGHPATSPHTNGPQDLRPSGTGVTFRSPVTSRARSISV 251
 DB 176 CSLGQRSHSLTLHNVSTGHPATSPHTNGPQDLRPSGTGVTFRSPVTSRARSISV 235
 QY 252 PSTPRRGATQKPKPKPK 269
 DB 236 PAPPRGACPPKPKPKPK 253

RESULT 3
 ID 096K25 PRELIMINARY; PRT; 177 AA.
 AC 096K25;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Hypothetical protein FLJ14935.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX PubMed=14702039; DOI=10.1038/ng1285;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Ohayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
 RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
 RA Sudo H., Hosoki T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takehashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
 RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,

RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
 RA Tanai H., Kimata M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,
 RA Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hottu T., Kusano J.,
 RA Kanemori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
 RA Togiya S., Komai K., Hara R., Takeuchi K., Arita M., Imose N.,
 RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
 RA Yoshioka Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
 RA Moriya S., Motiyama H., Satoh N., Takami S., Taraishi Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wabebe H.,
 RA Hisigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Saeki M.,
 RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shira Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs."
 RL Nat. Genet. 36:40-45 (2004).
 DR EMBL; AK027741; BAB55333.1; -
 SQ SEQUENCE 177 AA; 19050 MW; 834FDC0705AED0B CRC64;

Query Match 60.1%; Score 863; DB 2; Length 177;
 Best Local Similarity 99.4%; Pred. No. 4.8e-46;
 Matches 160; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTPVELAVSQMOTGLQHRGCRGVKARTSYVDETLFGSPAGTRPPDPDPVWEKA 60
 DB 1 MKTPVELAVSQMOTGLQHRGCRGVKARTSYVDETLFGSPAGTRPPDPDPVWEKA 60
 QY 61 NRTGCVGEAKKALGAKSCCTTPSGSTPTLPRKKNKRPISHTPSYCDSELRSSSE 120
 DB 61 NRTGCVGEAKKALGAKSCCTTPSGSTPTLPRKKNKRPISHTPSYCDSELRSSSE 120
 QY 121 GASRPAPMAKDAKRALALMTPTPTPGSHSPREAPL 161
 DB 121 GASRPAPMAKDAKRALALMTPTPTPGSHSPREAPL 161

RESULT 4
 ID 08S148 PRELIMINARY; PRT; 647 AA.
 AC 08S148;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE VsaA-like protein.
 GN Name=P0042A10.17;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzoae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=12447438; DOI=10.1038/nature01184;
 RA Saeki T., Matsunoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,
 RA Wu J., Nishimura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanemori H.,
 RA Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,
 RA Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,
 RA Hijishita S., Honda M., Ichikawa Y., Idenuma A., Iijima M., Ikeda M.,
 RA Ikeo M., Itoh S., Itoh T., Itoh Y., Iwabuuchi A., Kamiya K.,
 RA Karasawa M., Katsugiri S., Kikuta A., Kobayashi N., Kono I.,
 RA Machida K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y.,
 RA Nagasaka H., Nakashima M., Nakama Y., Nakamichi T., Nakamura M.,
 RA Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,
 RA Shimokawa T., Shomura A., Song J., Takasaki Y., Terasawa K., Teuji K.,
 RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yukawa K.,
 RA Zhong H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y.,

RA Waterston R.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Wilson R.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Wilson R.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RG WormBase Consortium;
 RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF000238; AAC48255.2;
 DR WormBase; WBGene0004157; W03D2.1.
 DR WormPep; W03D2.1a; CE26501.
 KM Prion.
 SQ SEQUENCE 524 AA; 51210 MW; 023D77EFD172CC5 CRC64;
 Query Match 10.1%; Score 144.5; DB 2; Length 524;
 Best Local Similarity 24.1%; Pred. No. 0.3;
 Matches 75; Conservative 18; Mismatches 97; Indels 121; Gaps 13;
 QY 13 QTLGLOHRCRGYKARTSYVDLTFGSPAGTPTPPDFPPVVEKANRTRGVGEKASK 72
 Db 219 QGFQGOOCTQNGFRKQKQ-----APAGSPPPP-----PP-----KQSP 253
 QY 73 ALGAKSGCETTPSGSTPTLTPRKKNKRYPISHTPSYCDESLFGSRSGAGAPRMKKG 132
 Db 254 PLASGSGSPPPPAGSPP--PPTGSPPPPTGSP-----PPPPAGG 293
 QY 133 DAAKLRLALWTPPTPRGS-----HSPRP-----REAPLRALHPA 167
 Db 294 SPPPRAGSPPPPPPRGSPPPTGSLPPRQAGSPPPACTGSPPPPRQKQAPERSPTG 353
 QY 168 GPSKTEP-----GPAADSQKLS-----MCG-----LHSSRPLKRG 198
 Db 354 SPTGSPPTGPRPGPKSSESESESRGCGRGGPRGGRPKKSSESESESRGPRGP 413
 QY 199 SHSLTLNLVSTGHPATSAHPHTNGQDLRPSTSGTFRSPLVTSARASVTSVSTPRRG 258
 Db 414 RRS-----PPTGSPPTGPRPGPRGSPTG-----SPTGLPSRQKQ 453
 QY 259 GATOKP--KPP 267
 Db 454 APEDRPTGSP 464

RESULT 7
 Q8M0G9 PRELIMINARY; PRT; 539 AA.
 ID Q8M0G9
 AC Q8M0G9;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Prion-like-(O/n-rich)-domain-bearing protein protein 75, isoform C.
 GN Name=pqn-75; ORFNames=W03D2.1;
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peleoderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=95069613; PubMed=9851916;
 RG WormBase Consortium;
 DE "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";

RL Science 282:2012-2018 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Rohlfing T.; Wohldmann P.;
 RT "The sequence of C. elegans cosmid W03D2.";
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Wilson R.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Wilson R.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RG WormBase Consortium;
 RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF000238; AAM97961.1;
 DR PIR; T28770; T28770.
 DR WormBase; WBGene0004157; W03D2.1.
 DR WormPep; W03D2.1c; CE14506.
 KM Prion.
 SQ SEQUENCE 539 AA; 52555 MW; D3D3CFD8EF42CDB9 CRC64;
 Query Match 10.1%; Score 144.5; DB 2; Length 539;
 Best Local Similarity 24.1%; Pred. No. 0.31;
 Matches 75; Conservative 18; Mismatches 97; Indels 121; Gaps 13;
 QY 13 QTLGLOHRCRGYKARTSYVDLTFGSPAGTPTPPDFPPVVEKANRTRGVGEKASK 72
 Db 214 QGFQGOOCTQNGFRKQKQ-----APAGSPPPP-----PP-----KQSP 268
 QY 73 ALGAKSGCETTPSGSTPTLTPRKKNKRYPISHTPSYCDESLFGSRSGAGAPRMKKG 132
 Db 269 PLASGSGSPPPPAGSPP--PPTGSPPPPTGSP-----PPPPAGG 308
 QY 133 DAAKLRLALWTPPTPRGS-----HSPRP-----REAPLRALHPA 167
 Db 309 SPPPRAGSPPPPPPRGSPPPTGSLPPRQAGSPPPACTGSPPPPRQKQAPERSPTG 368
 QY 168 GPSKTEP-----GPAADSQKLS-----MCG-----LHSSRPLKRG 198
 Db 369 SPTGSPPTGPRPGPKSSESESESRGCGRGGPRGGRPKKSSESESESRGPRGP 428
 QY 199 SHSLTLNLVSTGHPATSAHPHTNGQDLRPSTSGTFRSPLVTSARASVTSVSTPRRG 258
 Db 429 RRS-----PPTGSPPTGPRPGPRGSPTG-----SPTGLPSRQKQ 468
 QY 259 GATOKP--KPP 267
 Db 469 APEDRPTGSP 479

RESULT 8
 Q6NVE5 PRELIMINARY; PRT; 1067 AA.
 ID Q6NVE5
 AC Q6NVE5;
 DT 05-JUL-2004 (TREMBLrel. 27, Created)
 DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
 DT 05-JUN-2004 (TREMBLrel. 27, Last annotation update)
 DE Hypothetical protein LOC73072.
 GN Name=LOC73072;
 OS Mus musculus (Mouse).;

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien P.,
 RA Diatchenko L., Marusha K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loguclano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Villalón D.K., Muzny K.C., Hale S., Garcia A.M., Gay L.U., Hulyk S.W.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smalusz D.E., Schermer A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RC SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RA Straubeberg R.;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC068157; AA68157.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 1067 AA; 107615 MW; 561B9289E89FE824 CRC64;
 Query Match 10.0%; Score 144; DB 2; Length 1067;
 Best Local Similarity 26.7%; Pred. No. 0.66;
 Matches 74; Conservative 25; Mismatches 118; Indels 60; Gaps 13;
 QY 37 TLEGSP---AGTPTPPDPFPVWEKANR-----TRGVGKASALAKGSCETTP---SR 86
 DB 106 TIGAPLRASAGAPRSPANRPPASGKERAPMKTPEGGSISGPRASGIGARPPVQKR 165
 QY 87 GSTPTLTP--RKKNKRYPISTHPSYCDSELFSGRSEGSFGAP-----RWAKDAKRLA 139
 DB 166 LQPTPEPSARKGTPETPKNTLNSGTRRLVSDSLGTSIGASPAITRSRAPATB--- 222
 QY 140 LMTPTPTPGSHSPRPREAPLPAIHAPG--PSKTEPPPADSOKLSMGHSSRPPLKRL 198
 DB 223 -VGLPQAPAPARORPLTEAARKPGSSASRPSATELSPAFRRRSVAGGSL--OKPVSRL 279
 QY 199 SHSLTLHLNVSTG-----HFA-TSAPHTNTP--QDLRPSR 230
 DB 280 IPSATPOLSPBSRGVSPRTVPPRAPAHTSOLKSGQALHPTQTTVPKKNPSVQSLIPAS 339
 QY 231 SGVTFRSPVTSPRARSVISVSPSTPRGATQKRP 267
 DB 340 SLVTPTPP-----GASSVCGPDDPSQ--TTLRPSRP 368
 RESULT 9
 Q26963 PRELIMINARY; PRT; 964 AA.
 AC Q26963;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Trans-sialidase.
 GN Name=CTS-121;
 OS Trypanosoma cruzi.
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OC NCBI_TaxID=5693;
 RN [1]

RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=93010978; PubMed=1396577;
 RA Uemura H., Schenkman S., Nussenzweig V., Eichinger D.;
 RT "Only some members of a gene family in Trypanosoma cruzi encode
 RT proteins that express both trans-sialidase and neuraminidase
 RT activities.";
 RL EMBO J. 11:3837-3844(1992).
 RN [2]
 RC SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=97001677; PubMed=844669; DOI=10.1016/0166-6851(96)02634-5;
 RA Smith L.E., Uemura H., Eichinger D.;
 RT "Isolation and expression of an open reading frame encoding sialidase
 RT from Trypanosoma fangeli.";
 RL Mol. Biochem. Parasitol. 79:21-33(1996).
 DR HSP; Q26964; IMS3.
 DR GO:0004308; F:exo-alpha-sialidase activity; IEA.
 DR GO:0009405; P:phagocytosis; IEA.
 DR InterPro: IPR008985; ConA_like_1ec_g1.
 DR InterPro: IPR002860; Glyco_hydro_BNR.
 DR InterPro: IPR008377; Sialidase_trypan.
 DR Pfam: PF02012; BNR; 3.
 DR PRINTS: PRO1803; TCSIALIDASE.
 SQ SEQUENCE 964 AA; 101610 MW; 8F0600289FB847AD CRC64;
 Query Match 9.9%; Score 142.5; DB 2; Length 964;
 Best Local Similarity 26.8%; Pred. No. 0.74;
 Matches 84; Conservative 24; Mismatches 109; Indels 97; Gaps 19;
 QY 13 QTLGLHRCRGGR-----VKARTSYVDLTFGSPA----- 43
 DB 609 RTPDISHFYVGGRSDMPTISHVTNNVLLYNNROLNABEIRTLFLSODLIGTEAHMDS 668
 QY 44 -----GTPPTPPDPFPVWEKANRTRGV-----GKEASALAKGSCETTPS-----RG 87
 DB 669 SDGNMAHSTSTPD-----SSAHSSTPPADNGAHSTSTPADNGAHSTSTPPGNGAH 722
 QY 88 STPTLTPRKKNKRYPISTHPSYCDSELFSGRSEGSFGAPRNAKGDAKRLALMTPTPT 147
 DB 723 STP-LTPADNGAH--STPTPADSASHTSTPADNGA-----HSTPST 763
 QY 148 P--RGSHPRPREAPLPAIHAP--GSKTEPPPADSOKLSM-----GLHS--SRPLK 195
 DB 764 PADNGAHS-----TPTL--PADNGAHSSTPTP-ADSSAHSSTPPADNGAHSTSTPAD 814
 QY 196 RGLSHSLTLHLNVSTGHPATSAHTNGPOD--LRPSTGVTFRSPVTSPRARSVISVPS 253
 DB 815 NG-AHSTPTPADSSAHSSTPTPGDNGAHSTLTPADNG-AHSTPL--TPADSSAHSSTPS 870
 QY 254 TPRRGATQKRP 267
 DB 871 TPGDNGAHSTPSAP 884
 RESULT 10
 SYNI_CANFA STANDARD; PRT; 415 AA.
 ID SYNI_CANFA
 AC 062732;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Synapsin-1 (Synapsin I) (Fragment).
 GN Name=SYNI;
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OC NCBI_TaxID=9615;
 RN [1]
 RC SEQUENCE FROM N.A.

RA Shibuya H., Liu P.-C., O'Brien D.P., Johnson G.S.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Neuronal phosphoprotein that coats synaptic vesicles,
 binds to the cytoskeleton, and is believed to function in the
 regulation of neurotransmitter release (By similarity).
 CC -1- SUBCELLULAR LOCATION: Synapse.
 CC -1- SIMILARITY: Belongs to the synapsin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF049588; AAC05207.1; -.
 DR HSSP: P17599; 1AUX.
 DR InterPro: IPR001359; Synapsin.
 DR Pfam: PF02750; Synapsin C.1.
 DR PRINTS: PR01368; SYNAPSN.
 DR PROSITE: PS00415; SYNAPSN_1; PARTIAL.
 DR PROSITE: PS00416; SYNAPSN_2; PARTIAL.
 KW Actin-binding; Phosphorylation; Repeat; Synapse.
 FT NON_TER 1 1
 FT DOMAIN <1 148 C (Actin-binding and synaptic-vesicle
 binding).
 FT DOMAIN 149 383 D (Pro-rich linker).
 FT DOMAIN 384 >415 E.
 FT NON_TER 415 415
 SQ SEQUENCE 415 AA; 43388 MW; E988E2026FC5361B CRC64;

Query Match 9.7%; Score 140; DB 1; Length 415;
 Best Local Similarity 26.0%; Pred. No. 0.45;
 Matches 63; Conservative 17; Mismatches 108; Indels 54; Gaps 9;

QY 40 GSPAGTRTPDPDFPWEKANKTRG-----VGKASALAKSGCETTPRSSTP 90
 DB 182 GPAAQQR-PPQGGPFGPQGPQKQPLQAPPPQGGHISGLPPAG-SPLPRLSP 239
 QY 91 TLPRKKNKYPPISTHTPSYCDSELSFGSRSEGSFGAPPMKADAKLRLMTPTPTPR- 149
 DB 240 TSAPO-----QPVSGAQPLSGAQKQSRPVAGGAPAPPAAPRPPASPSPPQAGPPQARQ 294
 QY 150 -----GSHPPREAPLRAIHAPGPKTEPPGPAADSOKLSMGLSHSRPLKGLSHSLTHL 205
 DB 295 TSVSGAAPPKASGAPPSGQQRGPPQKPPGAPGPTROASQAG----- 336
 QY 206 NVPSGHPATAPAHNGPQDLRPSSTGVTFRSPVLTSSARSVSISVPTPRRGATQKPK 265
 DB 337 PMPRTGPTTQO-----RPSGGPAGR-PAKPKQLAKXPSQDVPSPATAG----- 381
 QY 266 PP 267
 DB 382 PP 383

RESULT 11
 Q6CT08 PRELIMINARY; PRT; 659 AA.
 AC Q6CT08;
 DT 25-OCT-2004 (TREMBLrel. 28, Created)
 DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
 DE Similar to trj095JC9 Sus scrofa Basic proline-rich protein.
 GN ORFNames=YALI0D261919;
 OS Yarrowia lipolytica CLIB99.
 OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Dipodascaceae; Yarrowia.
 RN NCB1_TaxID=284591;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLIB99;

RG Genolevures;
 RA Dulton B., Sherman D., Fischer G., Durrens P., Casaregola S.,
 RA Lafontaine I., de Montigny J., Marck C., Neveglie C., Talla E.,
 RA Goffard N., Frangeul L., Aigle M., Anthonard V., Babour A., Barbe V.,
 RA Barnay S., Blanchon S., Beckrich J.M., Beyne E., Bleykasten C.,
 RA Boierame A., Boyer J., Cattolico L., Confantolero F., de Darvar A.,
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
 RA Hantreay F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
 RA Nicoud J.M., Nikoliski M., Ozias S., Ozier-Kalogeropoulos O.,
 RA Pellenn S., Potier S., Richard G.F., Straub M.L., Suleau A.,
 RA Swenne D., Tekai F., Mesolowski-Leuvel M., Westhof E., Wirth B.,
 RA Zenitu-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
 RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,
 RA Wincker P., Souciet J.L.;
 RT "Genome evolution in yeasts."
 RN Nature 430:35-44(2004).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLIB99;
 RA Genoscope;
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: CR382130; CAG81510.1; -.
 DR InterPro: IPR003124; WH2.
 DR Pfam: PF02205; WH2.1.
 SQ SEQUENCE 659 AA; 62466 MW; 2CE5088BC9EF0908 CRC64;

Query Match 9.6%; Score 138.5; DB 2; Length 659;
 Best Local Similarity 26.0%; Pred. No. 0.88;
 Matches 70; Conservative 15; Mismatches 111; Indels 73; Gaps 11;

QY 42 PACGTRTPDPDFPWEKANKTRGVGKASALAKSGCETTPRSSTPILTPRKKNK 101
 DB 387 PGGAAPPIPGSAAAPAPAPAPAGAPAPAPSGA-----PTPPAPAPASAPAPAPAP 442
 QY 102 PISHTPSYCDSELSFGSRSEGSFGAPPMKADAKLRLMTPTPTPSGS-----PRP 156
 DB 443 PPSPPS-----TPGPMFGAP-MPKPAA---ASGAPPPPPGAAAPGAPAPAP 490
 QY 157 REAPLRAIHAPGPKTEBPAA-----DSQKL-----SMGLSHSRPLKGLSHSLTHLV 207
 DB 491 PAQPPSPGPGAPPPGPPAPPTDGFHSMILDDGSSGSHGAP-----P 537
 QY 208 PSTGHPATAPAHNGPQDLRPSSTGVTFRSPV-----TSRARSVS 248
 DB 538 PPPAPPSNGSHGAPPPPPNGVNRKRDVORTSLGSNNIRTLDTSAVTYAPRAVS 597
 QY 249 ISVPTPRRGATQKPKP-----WK 269
 DB 598 TPVSSSGGGGGS---KPPQIKIDSRWK 623

RESULT 12
 Q75MN6 PRELIMINARY; PRT; 2185 AA.
 AC Q75MN6;
 DT 05-JUL-2004 (TREMBLrel. 27, Created)
 DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
 DE Hypothetical protein MLL3 (Fragment).
 GN Name=MLL3;
 OS Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN NCB1_TaxID=9606;
 RX [1]
 RP MEDLINE=22737999; PubMed=12853948; DOI=10.1038/nature01782;
 RA Hillier L.W., Fulton R.S., Fulton L.A., Graves T.A., Pepin K.H.,
 RA Wagner-McPherson C., Layman D., Maas J., Jaeger S., Walker R.,
 RA Wylie K., Sekhon M., Becker M.C., O'Laughlin M.D., Schaller M.E.,
 RA Fewell G.A., Delehanuty K.D., Miner T.L., Nash W.B., Cordes M., Du H.,
 RA Sun H., Edwards J., Bradshaw-Cordum H., Ali J., Andrews S., Isaak A.,

RA Vanbrunt A., Nguyen C., Du F., Lamar B., Courtney L., Kalicki J.,
 RA Ozersky P., Bielicki L., Scott K., Holmes A., Harting R., Harris A.,
 RA Strong C.M., Hou S., Tomlinson C., Dauphin-Kohlberg S.,
 RA Kozlowicz-Reilly A., Leonard S., Rohlfing T., Rock S.M.,
 RA Latreille P., Miller N., Johnson D., Murray J., Strommatt C.,
 RA Lattin W.L.M., Abbott A., Minx P., Maupin R., Strommatt C.,
 RA Wendi M.C., Yang S.P., Schultz B.R., Wallis J.W., Spieth J.,
 RA Bieri T.A., Nelson J.O., Berkowicz N., Wohldmann P.B., Cook L.L.,
 RA Hickenbotham M.T., Eldred J., Williams D., Bedell J.A., Mardis E.R.,
 RA Clifton S.W., Chissee S.L., Marra M.A., Raymond C., Haugen E.,
 RA Gillett W., Zhou Y., James R., Phelps K., Iadamoto S., Bubbs K.,
 RA Simms E., Levy R., Clendenning J., Kaul R., Kent W.J., Pury T.S.,
 RA Baerbach R.A., Brent M.R., Keibler E., Flicek P., Bork P., Suyama M.,
 RA Bailey J.A., Portnoy M.E., Torrents D., Chinwalla A.T., Gish W.R.,
 RA Eddy S.R., McPherson J.D., Olson M.V., Eichler E.E., Green E.D.,
 RA Waterston R.H., Wilson R.K.,
 RT "the DNA sequence of human chromosome 7.",
 RL Nature 424:157-164(2003).

RM [2]
 RN SEQUENCE FROM N.A.

RA Bauer C., Gaije T., Nguyen C.,
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

RP [3]
 RN SEQUENCE FROM N.A.

RA Wilson R.,
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC006474; AAC00364.1; -
 DR GO; GO:0003677; F:DNA binding; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR InterPro; IPR000194; ATPase a/b-centre.

DR InterPro; IPR011011; FYVE_PHD_Znf.

DR InterPro; IPR000910; HMG_12_box.

DR InterPro; IPR001965; Znf_PHD.

DR Pfam; PF00505; HMG_box; 1.

DR Pfam; PF00628; PHD_3.

DR SMART; SM00398; HMG; 1.

DR SMART; SM00249; PHD; 3.

DR PROSITE; PS00152; ATPASE ALPHA BETA, UNKNOWN_1.

DR PROSITE; PS01359; ZF PHD 1; UNKNOWN_1.

DR PROSITE; PS50016; ZF PHD 2; 1.

KW Hypothetical protein_

FT NON TER 2185 2185

SO SEQUENCE 2185 AA; 239412 MW; 73A98266F35817E CRC64;

Query Match 9.5%; Score 136.5; DB 2; Length 2185;
 Best Local Similarity 24.0%; Pred. No. 4;
 Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARSYVDETLFGSPAGTR-----PTPPDFPPWVK-----ANRTGVGK 68

DB 994 SRLQNNETTANRPSVRLDSSSTNNNDPYAKPPTPRPVMTDQPKSLGLSRSPVSE 1053

QY 69 EASKALGAKGSC-----ETTPSGSTPLTLPRKKN-----YRPSHTPS 108

DB 1054 QIRKGLIAGTSHFTKPSRPADVFRQRIPDSYARPLTPADLDGPGGFTPMQPPPS 1113

QY 109 YCDSELFGRSEGCASGAPRMAGDAKRLALMTPTPTPGSH--SPRRAPPLRAIP 166

DB 1114 SQDP--YGSVSG-----ASRRLSVDYERPAL--TRPILDNSHNSQNDPYQGPLTTRHP 1164

QY 167 -----AGSK--TEG-----PAADSQKLSNGILHS-----SRPL 194

DB 1165 AVNESFAHSPRAFSQPGTISRPTSDPYQPGPTPRPVVDYSQSGSGTARSTNDPYSGP- 1223

QY 195 KRGLSHSLTHLWNPSTGHPATSAHPHTNGPDLRPSTGVTTPSPILVTSARSVSTSVPT 254

DB 1224 -----PGTPRTTVDYPSIQQTPTPTSTQTLTLVTPTVNHSDPYAHHPET 1270

QY 255 PRGGATQKPKPP 267

DB 1271 PRGGSIVPSYQPP 1283

RESULT 13

ML3_HUMAN STANDARD; PRT; 4911 AA.

ID Q9NE24; Q9NC02; Q9NDF6; Q9H9P4; Q9NR13; Q9P222; Q9UDR7;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Myeloid/lymphoid or mixed-lineage leukemia protein 3 homolog (Histone-lysine N-methyltransferase, H3 lysine-4 specific ML3) (EC 2.1.1.43)

DE (Homologous to ALR protein)

GN Name=ML3; Synonyms=HLR, KIAA1506;

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxId=9606;

RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Fetal thymus;
 RX MEDLINE=21888622; PubMed=11891048; DOI=10.1016/S0378-1119(02)00392-X;

RA Ruault M., Brun M.-E., Ventura M., Roizes G., De Sario A.,

RT "ML3, a new human member of the TRX/MLL gene family, maps to 7q36, a chromosome region frequently deleted in myeloid leukaemia.";

RL Gene 284:73-81(2002).

RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).

RC TISSUE=Cervical carcinoma;
 RX MEDLINE=21574953; PubMed=11718452;

RA Tan Y.C., Chow V.T.,

RT "Novel human HMLR (ML3) gene encodes a protein homologous to ALR and RT with leukemia and developmental defects.";

RL Cancer Detect. Prev. 25:454-469(2001).

RN [3]
 RP SEQUENCE FROM N.A.

RX MEDLINE=22737999; PubMed=12853948; DOI=10.1038/nature01782;

RA Haggler L.W., Fulton R.S., Fulton L.A., Graves T.A., Pepin K.H.,

RA Wagner-McPherson C., Layman D., Maas J., Jaeger S., Walker R.,

RA Wylie K., Sekhon M., Becker M.C., O'Laughlin M.D., Schaller M.E.,

RA Fowell G.A., Delehaunty K.D., Miner T.L., Nash W.E., Cordes M., Du H.,

RA Sun H., Edwards J., Bradshaw-Cordum H., Ali J., Andrews S., Isak A.,

RA Vanbrunt A., Nguyen C., Du F., Lamar B., Courtney L., Kalicki J.,

RA Ozersky P., Bielicki L., Scott K., Holmes A., Harting R., Harris A.,

RA Strong C.M., Hou S., Tomlinson C., Dauphin-Kohlberg S.,

RA Kozlowicz-Reilly A., Leonard S., Rohlfing T., Rock S.M.,

RA Latreille P., Miller N., Johnson D., Murray J., Strommatt C.,

RA Wendi M.C., Yang S.P., Schultz B.R., Wallis J.W., Spieth J.,

RA Bieri T.A., Nelson J.O., Berkowicz N., Wohldmann P.E., Cook L.L.,

RA Hickenbotham M.T., Eldred J., Williams D., Bedell J.A., Mardis E.R.,

RA Clifton S.W., Chissee S.L., Marra M.A., Raymond C., Haugen E.,
 RA Gillett W., Zhou Y., James R., Phelps K., Iadamoto S., Bubbs K.,
 RA Baerbach R.A., Brent M.R., Keibler E., Flicek P., Bork P., Suyama M.,
 RA Bailey J.A., Portnoy M.E., Torrents D., Chinwalla A.T., Gish W.R.,
 RA Eddy S.R., McPherson J.D., Olson M.V., Eichler E.E., Green E.D.,
 RA Waterston R.H., Wilson R.K.,
 RT "The DNA sequence of human chromosome 7.",
 RL Nature 424:157-164(2003).

RM [4]
 RN SEQUENCE OF 556-3865 FROM N.A. (ISOFORM 1).

RP TISSUE=Brain;
 RC TISSUE=Placenta;
 RX MEDLINE=20277482; PubMed=10819311;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.,
 RT "Prediction of the coding sequences of unidentified human genes. XVII.
 The complete sequences of 100 new cDNA clones from brain which code
 for large proteins in vitro.";
 RL DNA Res. 7:143-150(2000).

RM [5]
 RN SEQUENCE OF 3193-3865 AND 4460-4911 FROM N.A.
 RP TISSUE=Placenta;
 RC TISSUE=Placenta;
 RX PubMed=14702039; DOI=10.1038/ng1285;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,

RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA Nagahari K., Murakami K., Yasuda T., Iwanangi T., Wagatsuma M.,
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hirooka S., Chiba Y.,
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Houta T.,
RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,
RA Imose N., Mueshishima Y., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Teraishi Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukunaga Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujitani T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RT Nat. Genet. 36:40-45 (2004).
RL [6]
RN SEQUENCE OF 3879-4911 FROM N.A.
RP TISSUE-Testis;
RC Duesterhoeft A., Lauber J., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/Genbank/DBJ databases.
RN [7]
RP INTERACTION WITH ASC-2/NCOA6 CONTAINING COMPLEX (ISOFORM 2).
RC TISSUE=Cervical carcinoma;
RX MEDLINE=22371496; PubMed=12482968; DOI=10.1128/MCB.23.1.140-149.2003;
RA Gao Y.-H., Sohn Y.C., Kim D.-H., Kim S.-W., Kang M.-J., Jung D.-J.,
RA Kwak E., Barlev N.A., Berger S.L., Chow V.T., Roeder R.G.,
RA Azorin D.O., Meltzer P.S., Suh P.-G., Song E.J., Lee K.-J., Lee Y.C.,
RA Lee J.W.;
RT "Activating signal co-repressor 2 belongs to a novel steady-state
RT complex that contains a subset of trithorax group proteins.";
RL Mol. Cell Biol. 23:140-149 (2003).
CC -1- FUNCTION: Belongs to the ASC-2/NCOA6 complex (ASCOM), a
CC coactivator complex of nuclear receptors, involved in
CC transcriptional coactivation. MLJ3 may be a catalytic subunit of
CC this complex, which weakly methylates Lys-4 of histone H3. This is
CC a specific tag for epigenetic transcriptional activation. May be
CC involved in leukemogenesis and developmental disorder.
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + histone L-lysine =
CC S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine.
CC -1- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which
CC contains ASC-2/NCOA6, the retinoblastoma-binding protein Rb3-3/
CC RBP3, alpha- and beta-tubulins, the trithorax group proteins MLJ2
CC and MLJ3, and ASH2/ASCL2. Interacts with histone H3.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Name=1;
CC Event=Alternative splicing; Named isoforms=2;
CC Name=2;
CC IsoId=Q8NEZ4-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8NEZ4-2; Sequence=VSP_008561, VSP_008562;
CC -1- TISSUE SPECIFICITY: Highly expressed in testis and ovary, followed
CC by brain and liver. Also expressed in placenta, peripheral
CC blood, fetal thymus, heart, lung and kidney. Within brain,
CC expression was highest in hippocampus, caduate nucleus, and
CC substantia nigra. Not detected in skeletal muscle and fetal liver.
CC -1- DOMAIN: The SET domain interacts with histone H3 but not H2A, H2B
CC and H4, and may have a H3 lysine specific methylation activity.
CC -1- MISCELLANEOUS: Found in a critical region of chromosome 7, which
CC is commonly deleted in malignant myeloid disorders. Partial

CC duplication of the MLJ3 gene are found in the juxtaacentromeric
CC region of chromosomes 1, 2, 13 and 21. Juxtaacentromeric
CC -1- SIMILARITY: Belongs to the TRX/MLJ family.
CC -1- SIMILARITY: Contains 1 A-T hook DNA-binding repeat.
CC -1- SIMILARITY: Contains 1 DHHC-type zinc finger.
CC -1- SIMILARITY: Contains 6 PHD-type zinc fingers.
CC -1- SIMILARITY: Contains 1 post-SET domain.
CC -1- SIMILARITY: Contains 1 RING-type zinc finger.
CC -1- SIMILARITY: Contains 1 SET domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AV024361; AAK0583.1; -
CC EMBL: AF264750; AAF74766.2; -
CC EMBL: AC006017; AAD45822.1; -
CC EMBL: AC104692; -; NOT ANNOTATED CDS.
CC EMBL: AC005631; -; NOT ANNOTATED CDS.
CC EMBL: AB040939; BAA96030.2; -
CC EMBL: AK022687; BAB14179.1; -
CC EMBL: AK075113; BAC11409.1; -
CC EMBL: AL833924; CAD38780.1; -
CC HSSP: Q14839; 1MM2.
CC GeneW: HGNC:13726; MLJ3.
CC MIM: 606833; -
CC InterPro: IPR000637; A-T hook.
CC InterPro: IPR003889; FYRICH_C.
CC InterPro: IPR003888; FYRICH_N.
CC InterPro: IPR011011; FYVE_PHD_ZNF.
CC InterPro: IPR000910; HMG_12_box.
CC InterPro: IPR003616; PostSET.
CC InterPro: IPR001214; SET.
CC InterPro: IPR001594; Znf_DHHC.
CC InterPro: IPR001965; Znf_PHD.
CC InterPro: IPR001841; Znf_ring.
CC Pfam: PF02178; AT_hook; 1.
CC Pfam: PF05965; FYRICH; 1.
CC Pfam: PF05964; FYRICH; 1.
CC Pfam: PF00505; HMG_Box; 1.
CC Pfam: PF00628; PHD; 6.
CC Pfam: PF00856; SET; 1.
CC PROSITE: PS00354; HMG1_Y; 1.
CC PROSITE: PS50868; POST_SET; 1.
CC PROSITE: PS50280; SET; 1.
CC PROSITE: PS50216; ZF_DHHC; 1.
CC PROSITE: PS01359; ZF_PHD_1; 5.
CC PROSITE: PS50016; ZF_PHD_2; 6.
CC PROSITE: PS50089; ZF_RING_2; 1.
CC KW Activator; Alternative splicing; Chromatin regulator; Coiled coil;
CC DNA-binding; Methyltransferase; Nuclear protein; Polymorphism; Repeat;
CC Query Match 9.5%; Score 136.5; DB 1; Length 4911;
CC Best Local Similarity 24.0%; Pred. No. 9;
CC Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;
CC
CC Oy 29 ARTSYVDTEFLGSPAGR-----PTPPDPDPVAVK-----ANRTGVCK 68
CC Db 1933 SRPLQNMETNANRSPVRLDCSSSTTNNDPYAKPPDPVPTDQPKSLGLSRSPVSE 1992
CC Oy 69 EASKALGKSGC-----ETPSSRGSTPLPRKKNK-----YRPISHTPS 108
CC Db 1993 QTAGCPPLAAGSHFTPRSPADVFGQRIPDSARPLTPAPLPDSGPGPKTQWPPPS 2052
CC Oy 109 YCDSESLFGSRSEGSFQAPPAKDAKRALTLMTPTTPRGSH--SPREAPLRAIHP 166
CC Db 2053 SQDE--VGSVSG-----ASRLSVDPYERPAL--TPRIDNFSHNSQNDPYSQPLTPHP 2103

```

QY 167 -----AGPSK--TEPG-----PAADSOXKLSMGGLHS-----SRPL 194
DB 2104 AVNESFAHPSAFAFGQITSRPTSDPYSGPPGPRFVRVDSYSSSGSTASNTDPRYGP- 2162
QY 195 KRGLSHSLTHLNVSTGHPTASHTNGPODLRPSGSGVTFRSLVTSRAVSISVPT 254
DB 2163 -----GTRPRFTVDPYSQOPQTPRPSTQTDLFVTVTNQHRSHDPYAHPPGT 2209

QY 255 PRRGATOKPKP 267
DB 2210 PRGISTVYSGPP 2222

RESULT 14
Q9LLJ1 PRELIMINARY; PRT; 769 AA.
AC Q9LLJ1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DE CalK protein.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20379056; PubMed=10807915; DOI=10.1074/jbc.M002686200;
RA Pan J., Snell W.J.;
RT "Regulated targeting of a protein kinase into an intact flagellum. An
RT aureora/flip-like protein kinase translocates from the cell body into
RT the flagella during gamete activation in chlamydomonas."
RC J. Biol. Chem. 275:24106-24114(2000).
CL -1- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
DR HSPB; AF199021; AAF9501.1; -.
DR HSPB; O14965; 10L6
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro: IPR011009; Kinase like.
DR InterPro: IPR002290; Ser_thr_kinase.
DR InterPro: IPR008271; Ser_thr_pkinase.
DR Pfam: PF00069; Pkinase; 1.
DR ProDom: PD000001; Prot_kinase; 1.
DR SMART: SM00220; S_TKc; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 769 AA; 80249 MW; DB037FFCD4245101 CRC64;

Query Match 9.5%; Score 136; DB 2; Length 769;
Best Local Similarity 22.8%; Pred. No.1.5; Indels 98; Gaps 15;
Matches 75; Conservative 41; Mismatches 15;

QY 32 SYVDETLFG-----SPAGTRPTPPDDP-----WVERANR----- 62
DB 323 SEVEPQQLGFRSGNAGCGSPSGSKPPRLQVSSVAGGRLGHSSMSAMDKRYVD 382
QY 63 -----TRGVGKASAL-CAKSCCTTSRSGSTPLTLTRKKNK-----YRISH 105
DB 383 HILDVTLGMAAATLAKSVKNSVGDGDTTGGRPAGAGRVSRAGTANNSGTGAFLSR 442

QY 106 TRSYCDSELPFGSRSE-----GASFGAPRMAKGDAALRALIW-----TPPTPPGS 151
DB 443 NMSMGAOAVAAAAAAMASGSRGAHSGSDPKGRALTRBELTASLTAPLGGKSPVL 502
QY 152 -----HSPRPR-APLRAIHPACPSKTEPPADSOXKLSMGGLHSRPLKRG---LS 199
DB 503 ASFTAGRASPAPOQWAPL-----PTWAGKSGGGGGLGSRSSLSGATATAPSRGAELA 558
QY 200 HSLTHLNVSTGH-----PATSA-----PHTNG-----PODLRPSGSGVTFRSP 238

```

```

DB 559 HDLSRLRVDSAAAGCGGAACPSIASGAVPYSGKSSGSRISPPAPAPSTARGLTGSD 618
QY 239 LVTSPARSVISVSTPPRGATOKPKP 267
DB 619 LIPSRL-SQSFSPVSPPLAAGSNASAGTP 646

RESULT 15
Q7KSE8 PRELIMINARY; PRT; 2556 AA.
AC Q7KSE8;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, last annotation update)
DE CG7467-PC.
GN Name=osa; ORFNames=CG7467;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blasej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gaber G.L.,
RA Abail J.F., Aghayani A., An H.J., Andrews-Frankoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck Y., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Daventport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Idegam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheer F., Smith H.,
RA Shue B.C., Siden-Klamas I., Simpson M., Skupski M.P., Smit T.,
RA Spiter R., Spirdling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodgett M., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgeson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svitskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila

```

RL	melanogaster euchromatic genome sequence."	RT
RL	Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=22426070; PubMed=12537573;	
RA	Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,	
RA	Patel S., Frisoe E., Wheeler D.A., Lewis S.E., Rubin G.M.,	
RA	Ashburner M., Celisner S.E.,	
RT	"the transposable elements of the Drosophila melanogaster euchromatin:	
RT	a genomic perspective."	
RL	Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).	
RN	[4]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=22426069; PubMed=12537572;	
RA	Mitra S., Crosby M.A., Murgall C.J., Matthews B.B., Campbell K.S.,	
RA	Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,	
RA	Smith C.D., Tsuy J.L., Whitfield E.J., Bayraktaroglu I., Berman B.P.,	
RA	Bretencourt B.R., Celisner S.E., de Grey A.D., Drysdale R.A.,	
RA	Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,	
RA	Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,	
RA	Lewis S.E.;	
RT	"Annotation of the Drosophila melanogaster euchromatic genome: a	
RT	systematic review."	
RL	Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).	
RN	[5]	
RP	SEQUENCE FROM N.A.	
RX	FLYBase:	
RG	Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.	
RN	[6]	
RP	SEQUENCE FROM N.A.	
RG	FLYBase:	
RL	Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AE003718; AA6516.1; .	
DR	GO; GO:0005622; C:intracellular; IEA.	
DR	GO; GO:003677; F:DNA binding; IEA.	
DR	InterPro; IPR001606; ARID.	
DR	Pfam; PF01388; ARID. 1.	
DR	SMART; SM00501; BRIGHT. 1.	
SO	SEQUENCE 2556 AA; 268968 MW; 7F09E49CCB81D21E CRC64;	
Query Match	9.5%; Score 136; DB 2; Length 2556;	
Best Local Similarity	22.5%; Pred. No. 5;	
Matches	62; Conservative 30; Mismatches 119; Indels 64; Gaps 11	
OY	23 GGYRV-----KARTSYVDETLFGSPAGTRPTPDF-----DPWEKANTRGVG 67	
DB	633 GGYKMGCGQSGAGAGYPPQCPQYPPGVNPPRPPGAYATGPPPTSGA-----GAG 688	
OY	68 KEASALALANKSCETTPRSGTPTLTPRKKNKRRYISHTPSYCDSDLSFGSRBEGASFGAP 127	
DB	669 GANSPSPGAQAG--GYFGAG-----MPNHTGYPPYQWVPSPQQTVPGAGDGAAMVGNH 741	
OY	128 RMAKDAKALRALWLP-----DPTRGSHSPRPREALPAIHAPGSKTEPPADS 180	
DB	742 VQKKG-----TPRPVVGGRPPPGSSPPRLVYKHLQHKGGYGSPTPPQCP 791	
OY	181 QKLSTNG--GLHSRPL-----KRLSHSLTHLWVST-----GHPTSAPHT 220	
DB	792 QGYGNGPTGMPHGMPPHHNGPRPHGPGTNPMPSTPPQSQMLQGGCPQGGAGSGAPES 851	
OY	221 NGPDLRPSTGVTFRSPLVYSRAASVLSVSTP 255	
DB	852 GGPETH--SODNGISSGGPTGAAGMHAIVTSVVTGP 885	

Search completed: March 28, 2005, 08:53:15
Job time : 87 secs

This Page Blank (uspto)